

Report No. FAA-RD-81-21
FAA-CT-80-59

51
12

ANALYSIS OF DISCRETE WORD RECOGNITION

AD A105567

Dennis A. Steelman

FEDERAL AVIATION ADMINISTRATION TECHNICAL CENTER
Atlantic City Airport, N.J. 08405



INTERIM

JULY 1981

OCT 3 1981

A

Document is available to the U.S. public through
the National Technical Information Service,
Springfield, Virginia 22161.

MEC FILE COPY

Prepared for

U. S. DEPARTMENT OF TRANSPORTATION
FEDERAL AVIATION ADMINISTRATION
Systems Research & Development Service
Washington, D. C. 20590

034

Technical Report Documentation Page

1. Report No. 19 FAA-RD-81-21	2. Government Accession No. AD-A105 567	3. Recipient's Catalog No. 11 1261	
4. Title and Subtitle ANALYSIS OF DISCRETE WORD RECOGNITION	5. Report Date July 1981	6. Performing Organization Code	
7. Author(s) Dennis A. Steelman	8. Performing Organization Report No. 131-401-835	9. Work Unit No. (TRAIS)	
9. Performing Organization Name and Address Federal Aviation Administration Technical Center Atlantic City Airport, New Jersey 08405	10. Contractor Grant No. 14 FAA-CT-80-59	11. Type of Report and Period Covered Interim January 1980 to March 1980	
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Aviation Administration Systems Research and Development Service Washington, D. C. 20590	13. Sponsoring Agency Code	14.	
15. Supplementary Notes			
16. Abstract The Federal Aviation Administration Technical Center's Utterance Recognition Device (URD) was tested to determine its recognition rate and other pertinent operating characteristics for a vocabulary of 25 words. Audio input for the test was by means of standard voice grade telephone lines. No specific speaker training of the URD was performed prior to the test. Analysis of the resulting data base indicated that the 219 test subjects achieved an overall recognition rate of 85 percent. Computer simulation of subdividing the possible word choices, according to function-oriented subgroups, resulted in a 5 percent increase in the overall recognition rate.			
The results of this test will be used as reference for similar, future tests, using an expanded vocabulary to explore the possibility of using a device, such as the URD, as the input medium for direct user filing of flight plans over standard voice grade telephone lines.			
17. Key Words URD Recognition Rate Fast File Mass Weather Dissemination System Exploratory Engineering Model	18. Distribution Statement Document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 70	22. Price

PREFACE

The author wishes to acknowledge the invaluable assistance of Paul R. Quick, who provided the first and second choice word score distributions of appendix B, and Phillip Lui, who developed the Utterance Recognition Device (URD)/subject interface program. Thanks are also extended to Albert Schwartz and Howard Carlson for their data collection efforts.

It should be noted that Dialog Systems, Inc., manufacturer of the Utterance Recognition Device tested at the Federal Aviation Administration Technical Center, has been changed to Verbex. Interdata, the manufacturer of the host computer, has changed its name to Perkin Elmer Corporation.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
Purpose	1
Background	1
Discussion of Equipment	2
Theory of Operation	2
DISCUSSION	5
Test Procedure	5
Test Configuration	6
Test Results	8
Analysis	17
Summary of Results	21
CONCLUSIONS	21
RECOMMENDATIONS	22
REFERENCES	22
APPENDICES	
A - Computer Analyses (Including Glossary of Terms)	
B - Word Score Distribution	
C - Algorithm Explanation	

LIST OF ILLUSTRATIONS

Figure		Page
1	URD Vocabulary and Subgroups	1
2	Two-Pass Interpret Sequence	4
3	URD Raw Data Format	7
4	URD Raw Data (6 Subsets)	9
5	URD Test Installation	10
6	Computer Analysis of Total Data Base	12
7	Comparative Recognition Rate Graph	14
8	First Choice Word Distribution (2 Sheets)	15
9	Quantized Score Distribution	19
10	Idealized First and Second Choice Word Score Distribution	19
11	Proposed First and Second Choice Word Score Distribution For an Ideal Utterance Recognition Device	20

LIST OF TABLES

Table		Page
1	Data Base Distribution	6
2	URD Word Codes	8

INTRODUCTION

PURPOSE.

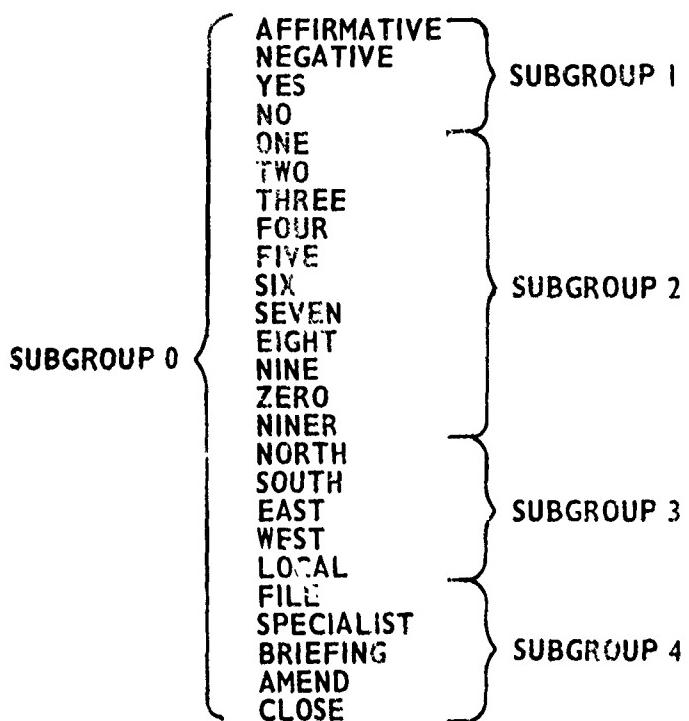
It is the purpose of this report to familiarize the reader with a preliminary, semiautomated test to determine the recognition rate of the Federal Aviation Administration (FAA) Technical Center's Utterance Recognition Device (URD) with a test vocabulary of 25 words. Figure 1 is a list of the vocabulary with subgroup restrictions.

The data collected in this test will be used to judge the effects of a proposed increase in the vocabulary size. An increase in vocabulary size will be necessary if the URD is to be used as the input medium for the direct filing of flight plans by utterance recognition.

BACKGROUND.

The FAA Technical Center's URD was originally acquired to explore the feasibility of using discrete word utterance recognition as the user control medium for the Mass Weather Dissemination System Exploratory Engineering Model (NPD 13-265). In this system, the URD was used to detect single word commands, over standard voice grade telephone lines, spoken by the caller. These command words were used to access the various functions of the engineering model. This application is reflected by the words which were selected for the original vocabulary which are given in figure 1.

One of the functions of the Mass Weather Dissemination System Exploratory Engineering Model was the Fast File. This function allowed the caller to file, amend or close a flight plan on one of two computer controlled cassette tape recorders. The actual recording of the flight plan information required no intervention by Flight Service Station (FSS) personnel. FSS



80-59-1

FIGURE 1. URD VOCABULARY AND SUBGROUPS

personnel intervention was, however, required to transcribe the resulting tapes several times per day for Service B transmission.

It has been proposed that the URD may be used as the input device for direct pilot flight plan filing by utterance recognition. This would serve to eliminate the labor intensive transcription phase of filing a flight plan.

Two previous URD tests have been conducted. In the first test a human monitor attempted to record URD responses to pilot utterances in order to determine the recognition rate of the URD. This method of testing proved cumbersome because of the difficulty in monitoring both subject and URD audio without affecting the electronic balance of the telephone connection.

The second URD test involved a human operator to simulate the URD in order to evaluate proposed human-URD communication protocols (reference 1).

DISCUSSION OF EQUIPMENT.

The FAA Technical Center URD under test is manufactured by Dialog Systems, Inc. The URD is asynchronously connected to and controlled by an Interdata 7/32 minicomputer which served as the host computer of the Mass Weather Dissemination System Exploratory Engineering Model.

The URD is a discrete word recognition machine; that is, it is capable of recognizing a single word at a time from a preprogrammed vocabulary.

The URD differs from the majority of voice recognition machines in three major aspects.

First, it is an untrained recognition device; that is, it is theoretically capable of recognizing, with equal success, any word in its vocabulary, regardless of speaker. This is an extremely desirable feature when one considers a system having thousands of potential users located remotely from the physical URD installation.

Second, the audio input to the URD is by means of standard voice grade telephone lines. This will serve to provide easy access to the user population. Standard voice grade telephone lines have a considerably smaller bandwidth than that of human speech. This complicates the recognition task since the upper and lower frequency components of the utterance are absent. Line noises and transients, inherent in switched communication systems of the magnitude of the standard phone system, must also be accounted for.

Finally, the URD is a multichannel recognition device, capable of handling up to eight different input channels

simultaneously. This capability is further expanded by multiplexing the eight independent input channels to 20 telephone lines by means of a cross point switching array controlled by the host computer.

The URD has limited speech capabilities. It is capable of saying any word in its recognition vocabulary as well as the phrases "Was that" and "Please repeat." These utterances are stored on an optical drum to provide relatively quick access. This vocabulary is used to seek verification of user utterances and provide confirmation to the caller.

The URD is controlled by the Interdata 7/32. The control functions of the 7/32 consist of connecting the caller to an available URD input channel, instructing the URD when to listen for a caller utterance, and then acting on the data returned by the URD.

At the present time, the URD vocabulary is configured into five subgroups as illustrated in figure 1. The utilization of subgroups allows the possible word choices to be restricted, thereby decreasing the possibility of the URD misunderstanding the utterance. For example, if a numeric input is expected, a direction-oriented word is obviously incorrect. The diversity of information required to file flight plans greatly restricts the use of subgroups to increase accuracy.

THEORY OF OPERATION.

When a speaker says a command word into his telephone instrument, the URD detects the utterance, processes it, and determines how closely the word corresponds to the stored reference templates of the subgroup of the vocabulary under consideration. Each possible word is then assigned a quality score which is inversely proportional to its probability of being the spoken word. The word having

the lowest quality score is referred to as the first choice word. Accordingly, the word having the next lowest score is referred to as the second choice word.

Three operational parameters exist to determine the quality of the first choice word. These parameters are termed GARBLE, VERIFICATION, and CONFUSION. It must be stressed, at this point, that these parameters in no way influence the score ranking of the vocabulary. All vocabulary elements are ranked prior to the application of the quality parameters.

The quality parameters may be modified by qualified personnel having access to the URD's command console. Default values for GARBLE, VERIFICATION, and CONFUSION exist and are, respectively, 3,560, 3,300, and 50.

Figure 2 is a flow diagram of how the URD employs the quality parameters to determine the probability of its first choice being correct. Upon being instructed by the host computer to listen for an utterance, the URD executes what is termed an interpret. An interpret consists of digitizing the audio input, analyzing it, and assigning quality scores to each word. If the quality score of the first choice word is greater than or equal to the value of GARBLE for that word, the URD will ask the speaker to repeat his utterance. If on the second attempt to recognize the utterance, the quality score is still greater than or equal to the value of GARBLE, the URD will ask the caller "Was that _____?" where _____ is the first choice word. If the user replies in the affirmative to this question, the first choice word code is transmitted to the host computer. A negative reply to the question will result in a code being transmitted to the host computer requesting that further corrective action be taken.

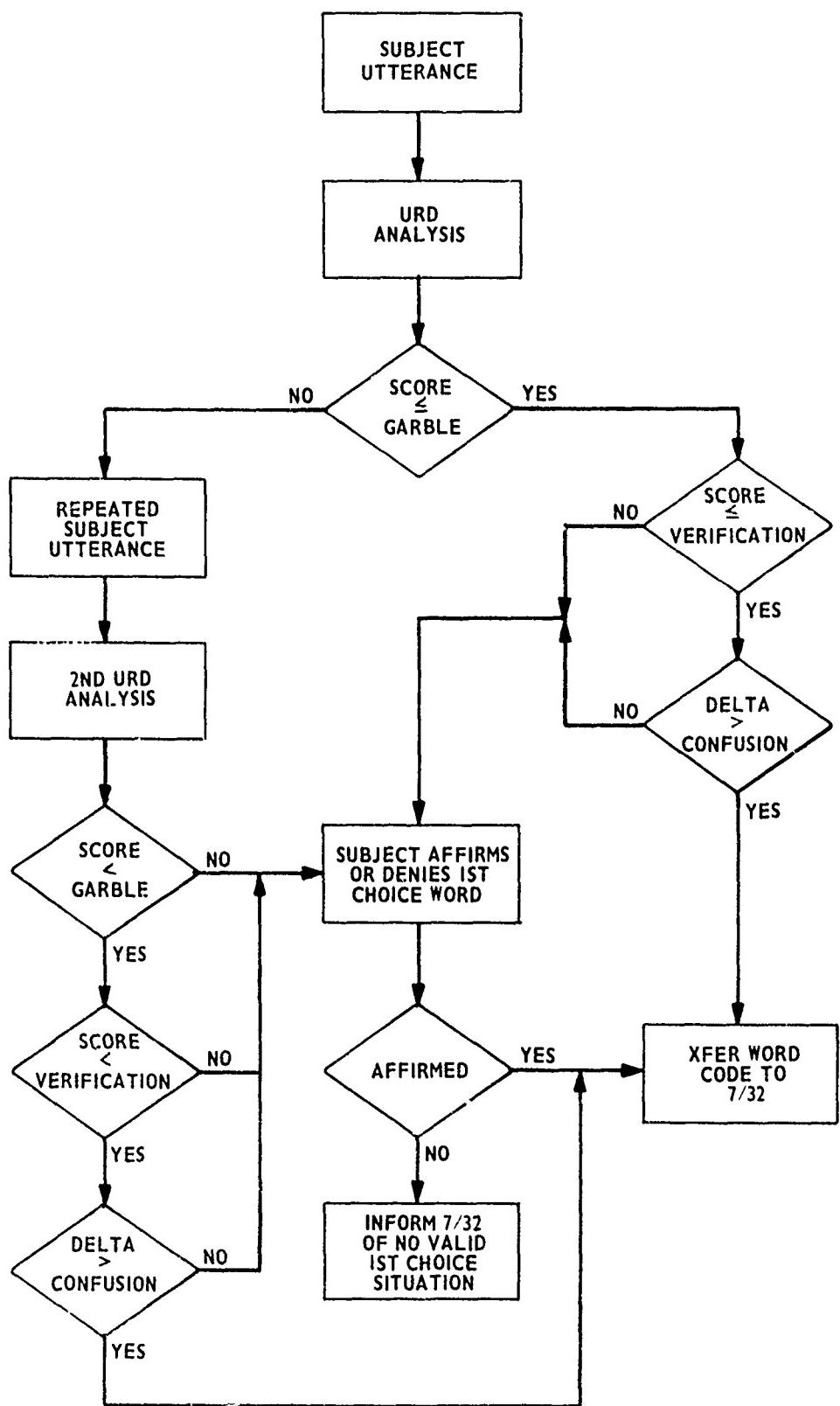
This sequence of confirming the first choice utterance by asking the user "Was that _____?" will be referred to as a WT sequence.

If the URD determines that the quality score is less than the value of GARBLE, it then checks to see if the quality score is less than the value of VERIFICATION for the first choice word. If the quality score is greater than or equal to VERIFICATION, a WT sequence is executed.

Assuming that the garble and verification tests have been passed successfully, the URD then checks the separation of the quality scores between the first and second choice words. If the separation is less than or equal to the value of CONFUSION for the first choice word, the URD will execute a WT sequence.

In the case where all three quality tests are passed, the URD assumes that a high probability exists that its first choice word is indeed the word spoken by the caller and transmits its code to the host computer. No confirmation is sought from the user in this situation.

Operating the URD with its default quality parameters will result in a comparatively large number of WT sequences being executed. Since this results in a second interpret being issued by the URD, this will be defined for the purpose of this report as a "two-pass system." If the quality parameters are modified so that confirmation is rarely sought of the caller, it shall be defined a "one-pass system." The same decisions regarding the quality of the first choice word are made in both systems. The difference lies strictly in the fact that the modified quality parameters virtually ensure that all quality tests will be passed.



80-59-2

FIGURE 2. TWO-PASS INTERPRET SEQUENCE

In its present form the URD has a vocabulary of 25 words. These words are divided into five subgroups as indicated in figure 1. Subgroup 0 is composed of all vocabulary elements. Subgroup 1 contains the AFFIRM/DENY words. Subgroup 2 contains the numbers ZERO through NINE, including NINER. Subgroup 3 is composed of direction-oriented words. Subgroup 4 contains those words which were used to access the specific functions of the Mass Weather Dissemination System Exploratory Engineering Model. Possible first choice words may be restricted to any one subgroup by the host computer when the interpret command is issued to the URD.

It should be noted that the actual method by which the URD determines its selection of the appropriate first choice word is proprietary to the URD's manufacturer.

DISCUSSION

TEST PROCEDURE.

The URD test was conducted using a diverse cross section of the FAA Technical Center's population composed of males, females, male pilots, and female pilots. Table 1 shows a numeric and percentile breakdown of the test population. Testing was performed at the subject's normal duty station so as to minimize any effect upon the subjects normally scheduled duties.

Two staff members of FSS laboratory functioned as a complementary data acquisition team. One member of the team contacted potential subjects in various locations at the Center. The other team member remained in the FSS laboratory.

The remote team member located a willing test subject, explained the test procedure, and notified the

in-house team member (via telephone) that a test sequence was about to begin. During the test sequence, the remote team member listened to the subjects utterances to ensure that each word was said in the proper sequence.

The in-house team member monitored the subject's audio as well as the URD's replies over an electronically isolated loudspeaker. This served to double-check that the subject spoke the test vocabulary in the proper sequence. The in-house team member also monitored the raw URD data on a cathode-ray tube (CRT) display to ensure that no serious anomalies occurred in the test data. Figure 3 is a detailed example of the data that appeared on the CRT. The single line of data presented in figure 3 shows that the subject accessed URD channel 0. The first choice word was AFFIRMATIVE (code 14) with a quality score of 3,266. The second choice word was THREE (code 3) with a quality score of 3,271. The amplitude of the utterance was 1,660. The utterance amplitude is a relative term and should be assigned no units by the reader. The reader should note that the utterance of figure 3 would have passed both the garble and verification tests but would have initiated a WT sequence due to insufficient score separation. Table 2 provides a list of the vocabulary elements and their word codes.

For the purposes of this test, the manufacturer's default quality parameters were modified to configure the URD as a single pass device. The new values for GARBLE, VERIFICATION, and CONFUSION were 3,560, 8,191, and 0, respectively. The virtual elimination of WT sequences served to greatly reduce the time required for a subject to complete a test run. Subject confusion was also reduced. In order to provide positive feedback to the subject, the URD was programmed to repeat the correct word from the vocabulary list, regardless of the interpretation of the first choice word.

TABLE 1. DATA BASE DISTRIBUTION

	<u>NUMBER OF SUBJECTS</u>	<u>TOTAL %</u>	<u>CORRECT %</u>
ALL SUBJECTS	219	100	85
MALES	198	90.4	85
FEMALES	21	9.6	82
PILOTS	59	26.9	86
MALE PILOTS	56	25.6	--
FEMALE PILOTS	3	1.4	--

In the test mode of operation, it would appear to a caller that the URD functioned at a 100-percent recognition rate. Later computer analysis of the data would be used to determine how well the URD had actually performed. All data obtained was stored in digital form on a disc file in the host computer. This file was then segmented into test length records with each record identified by subject name. An example of a segment of this data file is given in figure 4 and is read in the same way as figure 3. Subject identifiers have been removed from this report to conform with the Privacy Act of 1974 (Public Law 93-579).

It may be readily noted in figure 4 that subsets 2 and 3 achieved 100-percent first choice recognition. Despite this fact, in subset 2 both the word EIGHT (code 10) and the word BRIEFING would have induced WT sequences under normal operating conditions, due to quality scores in excess of VERIFICATION. It should be noted that the word EIGHT also would have failed to meet the confusion test if standard quality parameters were employed.

It may be noted in data subset 1, the word SIX (code 6) has been mistakenly interpreted as the word ZERO (code 0). This is actually a worst case mistake in that all quality conditions are met

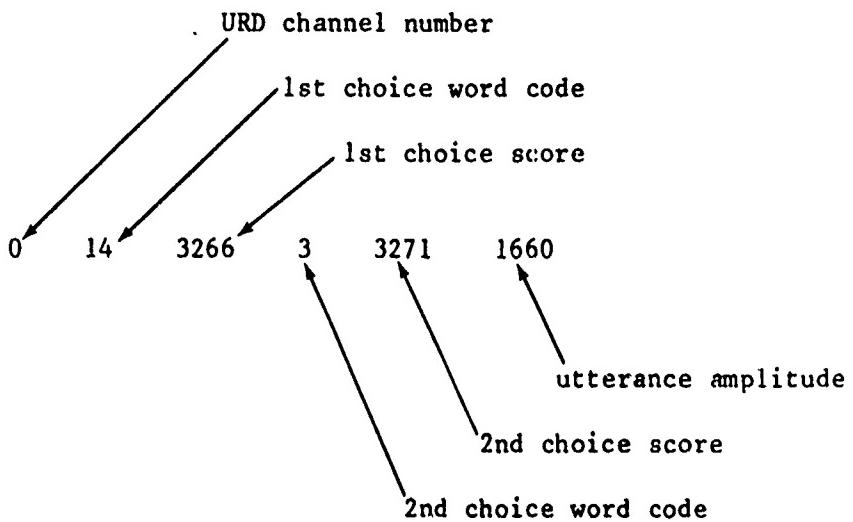
and the correct and incorrect word are both in the same subgroup. This will result in an incorrect word code being transmitted to the host computer without any confirmation being sought from the caller. These examples are indicated by -> in figure 4.

During the entire test, an audio record of all subject utterances was made on a standard 7-inch reel-to-reel tape recorder. It is proposed that this data will be used in the future to develop enhanced reference templates of the vocabulary elements.

TEST CONFIGURATION.

A diagram of the test installation is given in figure 5. The test subject dialed an outside line through the Center's switchboard using the telephone instrument at the site. This instrument may have been of the rotary dial type or Touch-Tone™. This call was then routed through the telephone switching office, located in Pleasantville, New Jersey, to a CDH D-mark located in the FSS laboratory, a total distance of approximately 8 air miles. At this point, the line was split between the URD and the digital speech output channel of the host computer.

The digital voice channel is used to provide the subject with an introductory preamble prior to the beginning



NOTES:

1. The channel number may be 0 through 7.
2. The first and second choice word codes have a direct correlation to a given vocabulary element. The word codes for each vocabulary element are given in table 2.
3. The first and second choice scores are used to determine the quality of the first choice word.
4. The utterance amplitude is a relative quantity and should be assigned no units.
5. The presence of a hyphen (-) between the 1st choice word code and its score indicates a time-out condition. In this case all data in the line, except the channel number, are invalid.

FIGURE 3. URD RAW DATA FORMAT

TABLE 2. URD WORD CODES

AFFIRMATIVE	14	ZERO.....	0
NEGATIVE.....	507	NINER.....	20
YES.....	13	NORTH.....	63
NO.....	12	SOUTH.....	64
ONE.....	1	EAST.....	65
TWO.....	2	WEST.....	66
THREE.....	3	LOCAL.....	67
FOUR.....	4	FILE.....	15
FIVE.....	5	SPECIALIST.....	21
SIX.....	6	BRIEFING.....	16
SEVEN.....	7	AMEND.....	16
EIGHT.....	10	CLOSE.....	2,663
NINE.....	11		

NOTE: In any case where a word code has more than four digits, the most significant digit is ignored. Example: 10013 and 13 are both the code for YES.

of a test sequence. It is also used to provide assistance to the user if the system determines that a severe recognition problem exists. This assistance is usually in the form of instructing the caller to proceed to the next word on the test list. The voice channel is only connected to the audio line when host computer audio output is required. At all other times it is gated out of the circuit by relay R1. These channels were designed to serve as the audio output devices for the Mass Weather Dissemination System Exploratory Engineering Model.

The audio input to the URD is also distributed through the URD to a voice-actuated tape recorder. This unit is used to acquire raw audio information to provide for future vocabulary enhancements. It also serves to drive the loudspeaker which is monitored by the in-house team member. This connection is provided by the URD manufacturer and does not affect the performance of the URD.

The URD is controlled by the host computer via asynchronous line 1 (AL1). The interpret commands and the URD's first choice word are passed on AL1.

The raw URD data, as shown in figure 3, is displayed on the CRT which is connected asynchronously to the URD. This displayed information is recorded by the host computer via asynchronous line 2 (AL2). AL2 is also utilized by the host computer to modify the default quality parameters to configure the URD as a single pass system. The data collected via AL2 is stored on one of the host computer's system discs so as to enable later, nonreal-time analysis of URD performance.

TEST RESULTS.

The data obtained in the URD test was subdivided into four subsets in order to determine if any grouping of participants had a significantly higher or lower recognition rate than the data base as a whole. The subgroups were

SUBSET 1											
0	14	2179	1766	1261	1766	2661	1766	22661	2515	2466	14
0	607	287	21	65	507	524	67	582	2969	507	3247
0	11	246	64	177	11	168	7	349	2867	11	2327
0	12	655	64	261	2704	16012	1772	2661	1644	12	3054
0	1	115	20661	197	3561	16001	111	20661	514	2266	54
0	641	641	113	145	113	10661	115	19666	164	2414	6
0	10661	157	26661	47	145	10661	115	115	1452	2266	9
0	697	16061	16061	16061	16061	16061	4	122	92	275	6
0	16061	113	46661	146	146	16061	5	1314	15	1644	6
0	50661	116	116	116	116	16061	6	684	67	2661	5
0	19661	21	116	116	116	10661	7	699	65	2266	6
0	19	2188	20665	1266	1266	10661	12	10661	126	10661	7
0	20661	11	1164	1541	1541	10661	11	10661	164	275	9
0	2951	46661	4667	16061	16061	16061	2	2969	2295	2975	9
0	10	176	17	17	17	26661	1	2969	16664	1422	1616
0	46664	62	62	144	124	1252	1	1817	15	3175	15
0	4	215	50666	1274	1274	62	2264	46664	175	2661	1275
0	65	50669	461	17	1626	64	2064	20657	1422	1644	64
0	66	185	256	67	272	65	2175	50667	126	275	65
0	67	215	50661	175	272	65	2168	50661	1272	4764	66
0	15	118	50661	274	267	67	171	50661	117	5464	67
0	21	261	261	147	177	15	1687	64	15	6819	65
0	16	17	17	417	1649	21	2691	26915	1422	1422	21
0	17	162	14	177	1272	16	26626	1644	2563	16	3179
0	2661	14	454	1472	1472	17	267	65	2425	17	2218
0	1166	14	14	14	14	2661	1259	2661	2661	2661	2179
											1002

SUBSET 2											
0	14	2179	1766	1261	1766	2661	1766	22661	2515	2466	14
0	607	287	21	65	507	524	67	582	2969	507	3247
0	11	246	64	177	11	168	7	349	2867	11	2327
0	12	655	64	261	2704	16012	1772	2661	1644	12	3054
0	1	115	20661	197	3561	16001	111	20661	514	2266	54
0	641	641	113	145	113	10661	115	19666	164	2414	6
0	10661	157	26661	47	145	10661	115	115	1452	2266	9
0	697	16061	16061	16061	16061	16061	4	122	92	275	6
0	16061	113	46661	146	146	16061	5	1314	15	1644	6
0	50661	116	116	116	116	16061	6	684	67	2661	5
0	19661	21	116	116	116	10661	7	699	65	2266	6
0	19	2188	20665	1266	1266	10661	12	10661	164	275	9
0	20661	11	1164	1541	1541	10661	11	10661	164	275	9
0	2951	46661	4667	16061	16061	16061	2	2969	16664	1422	1616
0	10	176	17	177	177	1772	1	1817	15	3175	15
0	46664	62	62	144	124	1252	1	1644	164	1644	6
0	4	215	50666	1274	1274	62	2064	20657	1422	1644	6
0	65	50669	461	17	1626	65	2175	50667	126	275	65
0	66	185	256	67	272	65	2168	50661	1272	4764	66
0	67	215	50661	274	267	65	171	50661	117	5464	67
0	15	118	50661	274	267	15	1687	64	15	6819	65
0	21	261	261	147	177	15	16915	46666	2275	2275	2
0	16	17	17	417	1649	15	16911	16911	1644	1644	2
0	2661	14	14	14	14	10661	6	684	67	2661	5
0	19661	112	116	116	116	10661	6	2661	2661	2661	112
0	641	641	113	145	113	10661	6	19661	164	275	6
0	10661	157	26661	47	145	10661	6	10661	164	275	9
0	697	16061	16061	16061	16061	16061	4	122	92	275	6
0	16061	113	46661	146	146	16061	5	1314	15	1644	6
0	50661	116	116	116	116	16061	6	684	67	2661	5
0	19661	21	116	1541	1541	10661	12	10661	164	275	9
0	19	215	50666	1274	1274	10661	12	10661	164	275	9
0	20661	11	1164	1541	1541	10661	11	10661	164	275	9
0	2951	46661	4667	16061	16061	16061	2	2969	16664	1422	1616
0	10	176	17	177	177	1772	1	1817	15	3175	15
0	46664	62	62	144	124	1252	1	1644	164	1644	6
0	4	215	50669	461	17	1626	5	10661	164	275	6
0	65	185	256	67	272	5	2114	50661	126	275	6
0	66	215	50661	274	267	5	1644	50661	1272	4764	66
0	67	118	50661	274	267	5	1644	50661	117	5464	67
0	21	261	261	147	177	5	16915	16915	1644	1644	2
0	16	17	17	417	1649	5	16911	16911	1644	1644	2
0	2661	14	14	14	14	10661	5	684	67	2661	5
0	19661	112	116	116	116	10661	6	19661	164	275	6
0	641	641	113	145	113	10661	6	10661	164	275	9
0	10661	157	26661	47	145	10661	6	10661	164	275	9
0	697	16061	16061	16061	16061	16061	4	122	92	275	6
0	16061	113	46661	146	146	16061	5	1314	15	1644	6
0	50661	116	116	116	116	16061	6	684	67	2661	5
0	19661	21	116	1541	1541	10661	12	10661	164	275	9
0	19	215	50666	1274	1274	10661	12	10661	164	275	9
0	20661	11	1164	1541	1541	10661	11	10661	164	275	9
0	2951	46661	4667	16061	16061	16061	2	2969	16664	1422	1616
0	10	176	17	177	177	1772	1	1817	15	3175	15
0	46664	62	62	144	124	1252	1	1644	164	1644	6
0	4	215	50669	461	17	1626	5	10661	164	275	6
0	65	185	256	67	272	5	2114	50661	126	275	6
0	66	215	50661	274	267	5	1644	50661	1272	4764	66
0	67	118	50661	274	267	5	1644	50661	117	5464	67
0	21	261	261	147	177	5	16915	16915	1644	1644	2
0	16	17	17	417	1649	5	16911	16911	1644	1644	2
0	2661	14	14	14	14	10661	5	684	67	2661	5
0	19661	112	116	116	116	10661	6	19661	164	275	6
0	641	641	113	145	113	10661	6	10661	164	275	9
0	10661	157	26661	47	145	10661	6	10661	164	275	9
0	697	16061	16061	16061	16061	16061	4	122	92	275	6
0	16061	113	46661	146	146	16061	5	1314	15	1644	6
0	50661	116	116	116	116	16061	6	684	67	2661	5
0	19661	21	116	1541	1541	10661	12	10661	164	275	9
0	19	215	50666	1274	1274	10661	12	10661	164	275	9
0	20661	11	1164	1541	1541	10661	11	10661	164	275	9
0	2951	46661	4667	16061	16061	16061	2	2969	16664	1422	1616
0	10	176	17	177	177	1772	1	1817	15	3175	15
0	46664	62	62	144	124	1252	1	1644	164	1644	6
0	4	215	50669	461	17	1626	5	10661	164	275	6
0	65	185	256	67	272	5	2114	50661	126	275	6
0	66	215	50661	274	267	5	1644	50661	1272	4764	66
0	67	118	50661	274	267	5</td					

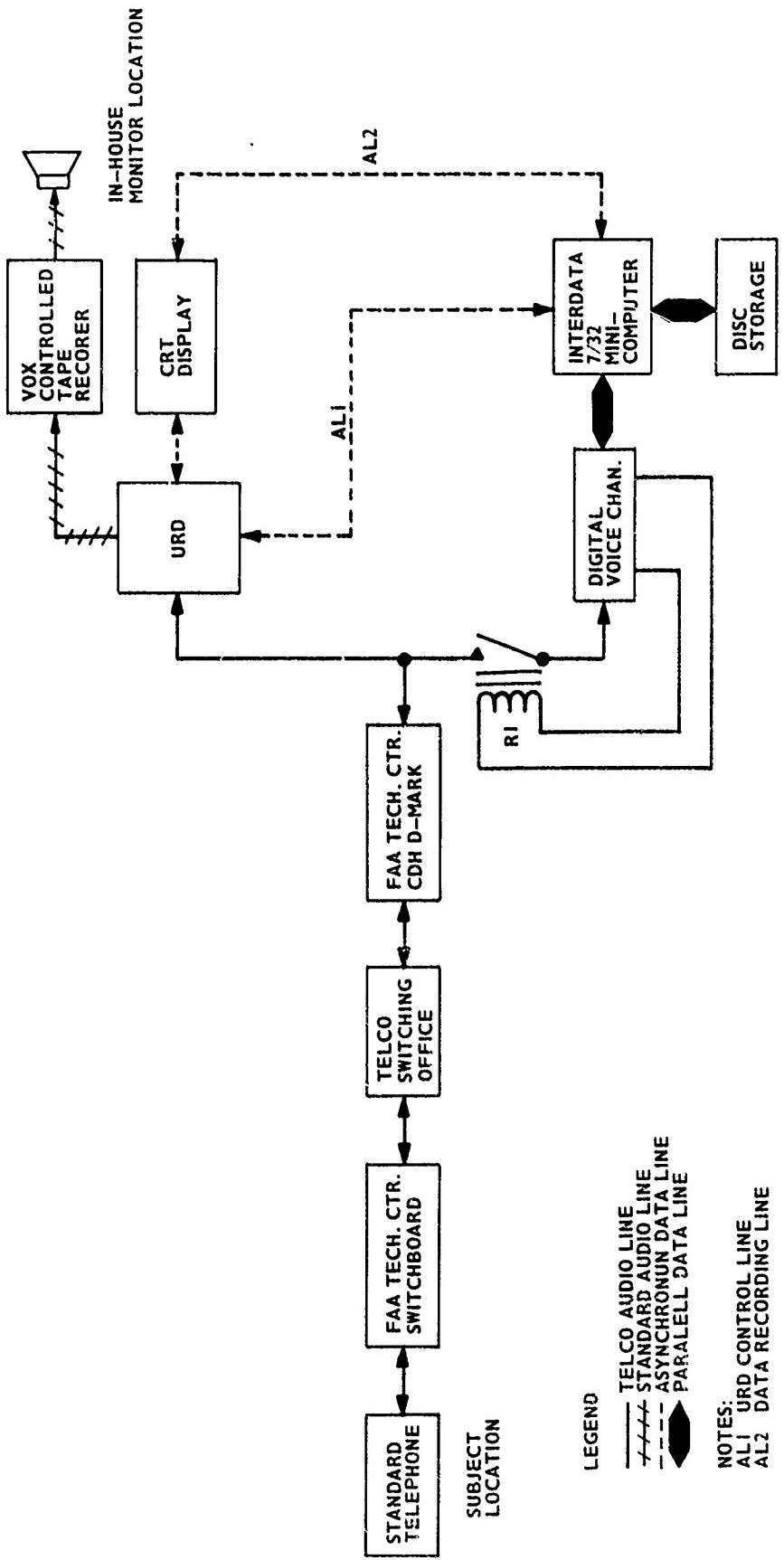


FIGURE 5. URD TEST INSTALLATION

80-59-5

composed of (1) pilots, (2) males only, (3) females only, and (4) the entire data base consisting of 219 test subjects.

Table 1 shows that the entire data base had a recognition rate of 85 percent. Ranking the other three subsets of the data in descending order, in terms of recognition rates, pilots were first, males second, and females lowest in overall recognition rate of the subsets. The percentile figures presented in table 1 were obtained by computer analysis of the raw data obtained in the URD test. An example of this analysis for the entire data base is given in figure 6. The value of employing computer analysis lies in the fact that large amounts of data regarding URD performance under various conditions, such as modified quality parameters and segregation of the test population, may be obtained with only one data collection exercise for a given vocabulary.

The first page of the computer analysis for the entire data base, which is given in figure 6, contains considerable information regarding both the URD's performance and the simulated conditions of the experimental run. The name of the data file under consideration is found in the upper left hand corner of figure 6. In this case the file is ALL URD which represents the total data base. The quality parameters for the analysis run are given next. In the case of figure 6, the default parameters, 3,560, 3,300, and 50 were selected. The remainder of figure 6 is devoted to the actual analysis of the data.

The word AFFIRMATIVE will be used as an example of how to read the analysis presented in figure 6. Proceeding from left to right, it may be seen that the word AFFIRMATIVE was correctly selected as the first choice word 177 times (RIGHT). A word other than AFFIRMATIVE was selected incorrectly as the first

choice 43 times (WRONG). Using the quality parameters given previously, the URD would have sought confirmation of an utterance by initiating a WT sequence 148 times (WT). A garble condition would have caused the URD to ask the caller to "Please repeat!" once (PR). The total number of valid interprets for a word is equal to the sum of RIGHT + WRONG.

The URD failed to detect sufficient audio to consider an utterance valid in six cases. This situation is referred to as a time-out (TO). In 21 of the cases in which the first choice word was incorrect, the second choice word was the correct word (SEC). The word AFFIRMATIVE is in subgroup 1. In 42 of the 43 incorrect first choices, the subgroup of the first choice word was incorrect (SGW).

The word AFFIRMATIVE was correctly chosen as the first choice word in 80 percent of the cases (%R). In 39 of the cases in which the URD sought confirmation of an utterance via a WT sequence, the first choice word was incorrect (WWT).

When the word AFFIRMATIVE was correctly recognized as the first choice word, it had an average quality score of 3,298 (MEAN) with a standard deviation from this average of 58 (STDM). The average separation between the score of the word AFFIRMATIVE and the second choice word's quality score was 82 (DELTA). The standard deviation from DELTA was 50 (STDD).

In those cases where AFFIRMATIVE was correctly recognized as the word spoken, the average utterance amplitude was 1,938 (AMPR). The standard deviation from AMPR was 1,277 (STDAM). In those cases where another vocabulary element was substituted in place of the word AFFIRMATIVE, the average utterance amplitude was 1,390 (AMPW). The standard deviation from AMPW was 1,030 (STDW).

ALL	UP	WORD	AFFIRMATIVE	NEUTER	NEGATIVE	FILE	REPORT	CONFUSION
GRABBLE	1550	1550	1550	1550	1550	1550	1550	50
AFFIRMATIVE	1550	1550	1550	1550	1550	1550	1550	50
NEGATIVE	1550	1550	1550	1550	1550	1550	1550	50
YES	1550	1550	1550	1550	1550	1550	1550	50
NO	1550	1550	1550	1550	1550	1550	1550	50
ONE	1550	1550	1550	1550	1550	1550	1550	50
TWO	1550	1550	1550	1550	1550	1550	1550	50
THREE	1550	1550	1550	1550	1550	1550	1550	50
FOUR	1550	1550	1550	1550	1550	1550	1550	50
FIVE	1550	1550	1550	1550	1550	1550	1550	50
SIX	1550	1550	1550	1550	1550	1550	1550	50
SEVEN	1550	1550	1550	1550	1550	1550	1550	50
EIGHT	1550	1550	1550	1550	1550	1550	1550	50
NINE	1550	1550	1550	1550	1550	1550	1550	50
ZERO	1550	1550	1550	1550	1550	1550	1550	50
NORTH	1550	1550	1550	1550	1550	1550	1550	50
WEST	1550	1550	1550	1550	1550	1550	1550	50
LOCAL	1550	1550	1550	1550	1550	1550	1550	50
FILE	1550	1550	1550	1550	1550	1550	1550	50
SPECIALIST	1550	1550	1550	1550	1550	1550	1550	50
BELIEFING	1550	1550	1550	1550	1550	1550	1550	50
AMEND	1550	1550	1550	1550	1550	1550	1550	50
CLOSE	1550	1550	1550	1550	1550	1550	1550	50
WORD	1550	1550	1550	1550	1550	1550	1550	50
AFFIRMATIVE	1550	1550	1550	1550	1550	1550	1550	50
NEUTER	1550	1550	1550	1550	1550	1550	1550	50
NEGATIVE	1550	1550	1550	1550	1550	1550	1550	50
YES	1550	1550	1550	1550	1550	1550	1550	50
NO	1550	1550	1550	1550	1550	1550	1550	50
ONE	1550	1550	1550	1550	1550	1550	1550	50
TWO	1550	1550	1550	1550	1550	1550	1550	50
THREE	1550	1550	1550	1550	1550	1550	1550	50
FOUR	1550	1550	1550	1550	1550	1550	1550	50
FIVE	1550	1550	1550	1550	1550	1550	1550	50
SIX	1550	1550	1550	1550	1550	1550	1550	50
SEVEN	1550	1550	1550	1550	1550	1550	1550	50
EIGHT	1550	1550	1550	1550	1550	1550	1550	50
NINE	1550	1550	1550	1550	1550	1550	1550	50
ZERO	1550	1550	1550	1550	1550	1550	1550	50
NORTH	1550	1550	1550	1550	1550	1550	1550	50
WEST	1550	1550	1550	1550	1550	1550	1550	50
LOCAL	1550	1550	1550	1550	1550	1550	1550	50
FILE	1550	1550	1550	1550	1550	1550	1550	50
SPECIALIST	1550	1550	1550	1550	1550	1550	1550	50
BELIEFING	1550	1550	1550	1550	1550	1550	1550	50
AMEND	1550	1550	1550	1550	1550	1550	1550	50
CLOSE	1550	1550	1550	1550	1550	1550	1550	50
TOTAL	1550	1550	1550	1550	1550	1550	1550	50
TOTAL	4576	847	847	847	847	847	847	548
TOTAL	548	85	85	85	85	85	85	85
TOTAL	PERCENT CORRECT	PERCENT INCORRECT	PERCENT THAT					

FIGURE 6. COMPUTER ANALYSIS OF TOTAL DATA BASE

If the URD had been issuing subgroup restricted interprets in the course of the test, the recognition rate would have been greater than or equal to 90 percent (M%R). Finally, 41 of the 148 WT sequences that occurred on the word AFFIRMATIVE were due to confusion situations (CON).

In those cases, where it is applicable, total figures are given for all vocabulary elements. These figures are given in the row labeled TOTALS. A more detailed explanation of the column headers is given in appendix A.

A brief summary of the test run is given in the lower left-hand corner of figure 6. Appendix A contains complete runs for each data subset and an analysis of the entire date base, using both standard and modified quality parameters.

A graphical representation of the comparative recognition rate for each element in the vocabulary is given in figure 7. The reader's attention is directed to the word SOUTH. Females achieved a markedly lower recognition rate (39 percent) for this word than did the predominately male subgroups. Conversely, females achieved a much higher recognition rate (100 percent) for the word TWO. When the data base as a whole is considered, as presented in table 1, females achieved a recognition rate of only 3 percent less than the predominately male total data base. The reader should be aware from table 1 that females composed only a small percentage of the data base (9.6 percent). No separation between male and female pilots was made due to the small number (3) of female pilots readily available for test purposes.

Figure 8 is a first choice word distribution analysis for the entire data base. It shows how many times each vocabulary element was selected as the first choice word when a particular

word was expected. Referring to figure 8, the word at the top of each column is the correct first choice word. The figures in each column indicate how many times the associated word was selected as the first choice. Taking the word AFFIRMATIVE as an example, it is readily noted that AFFIRMATIVE was selected correctly as the first choice word 177 times. This corresponds to the value of RIGHT for the word AFFIRMATIVE in figure 6. In the same column, it is shown that the URD incorrectly selected the vocabulary element CLOSE as the first choice word 13 times. If all entries in the column except those corresponding to the correct first choice word are totaled, the sum will equal the value of WRONG for the given vocabulary element. In the case of the word AFFIRMATIVE, this total is 43 which is equal to the value of WRONG for AFFIRMATIVE in figure 6.

The reader's attention is directed to the column for the word NINE. This word was mistaken for FIVE 21 times and NINER 10 times. This is an example of a worst case situation in which the words most commonly mistaken for the correct utterance are members of the same vocabulary subgroup as the correct first choice word.

The final column in figure 8 is labeled TOTALS. The figures in this column represent the total number of times each word in the vocabulary was selected as the first choice word. These values may be considered as the sum across a row of all 25 columns. For an example, the word AFFIRMATIVE was selected as the first choice word 199 times. This column shows that, for the entire data base, the word most commonly selected as the first choice was ONE. The least commonly selected first choice word was NINE.

Distribution analyses for all subsets are found in appendix A.

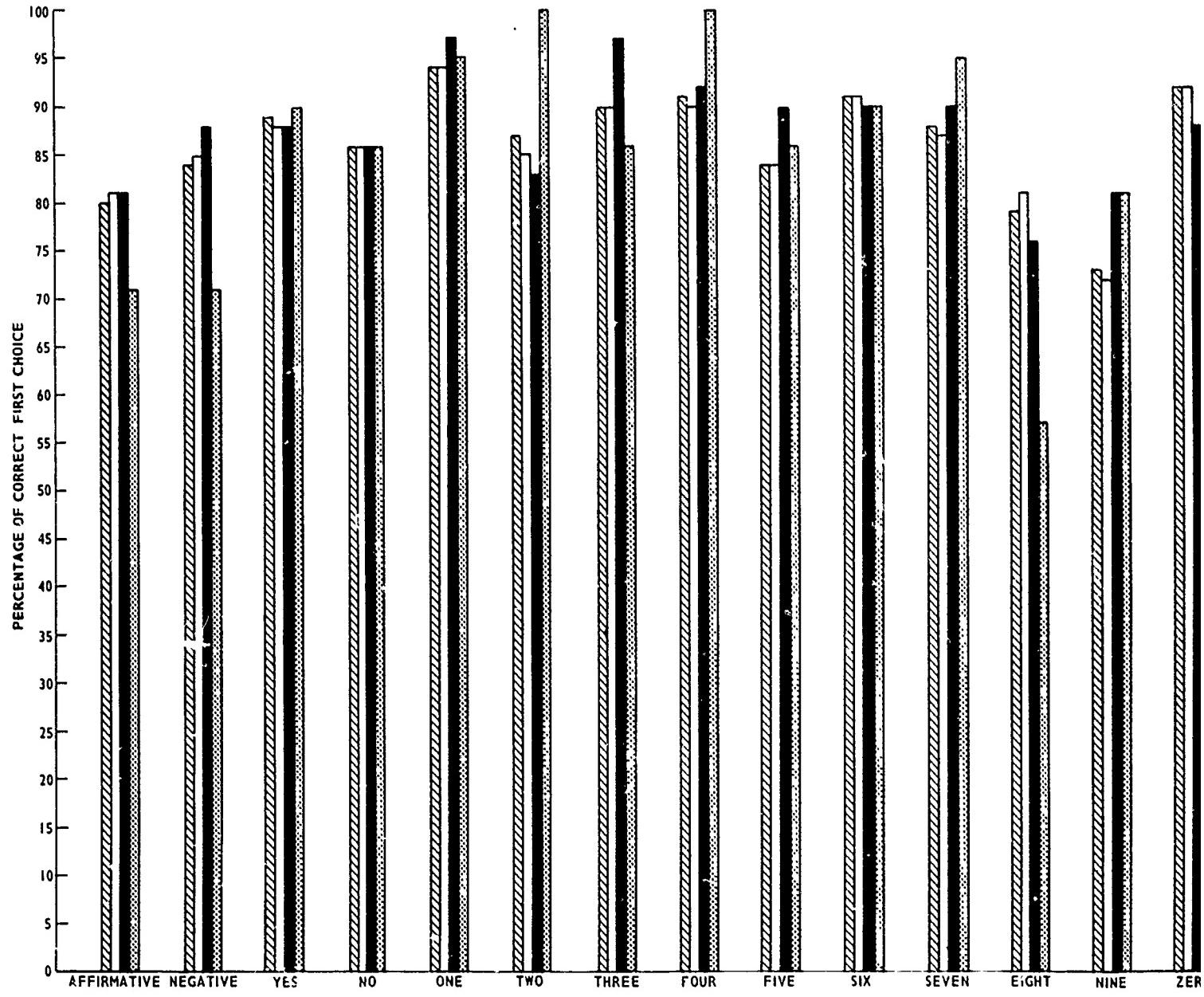
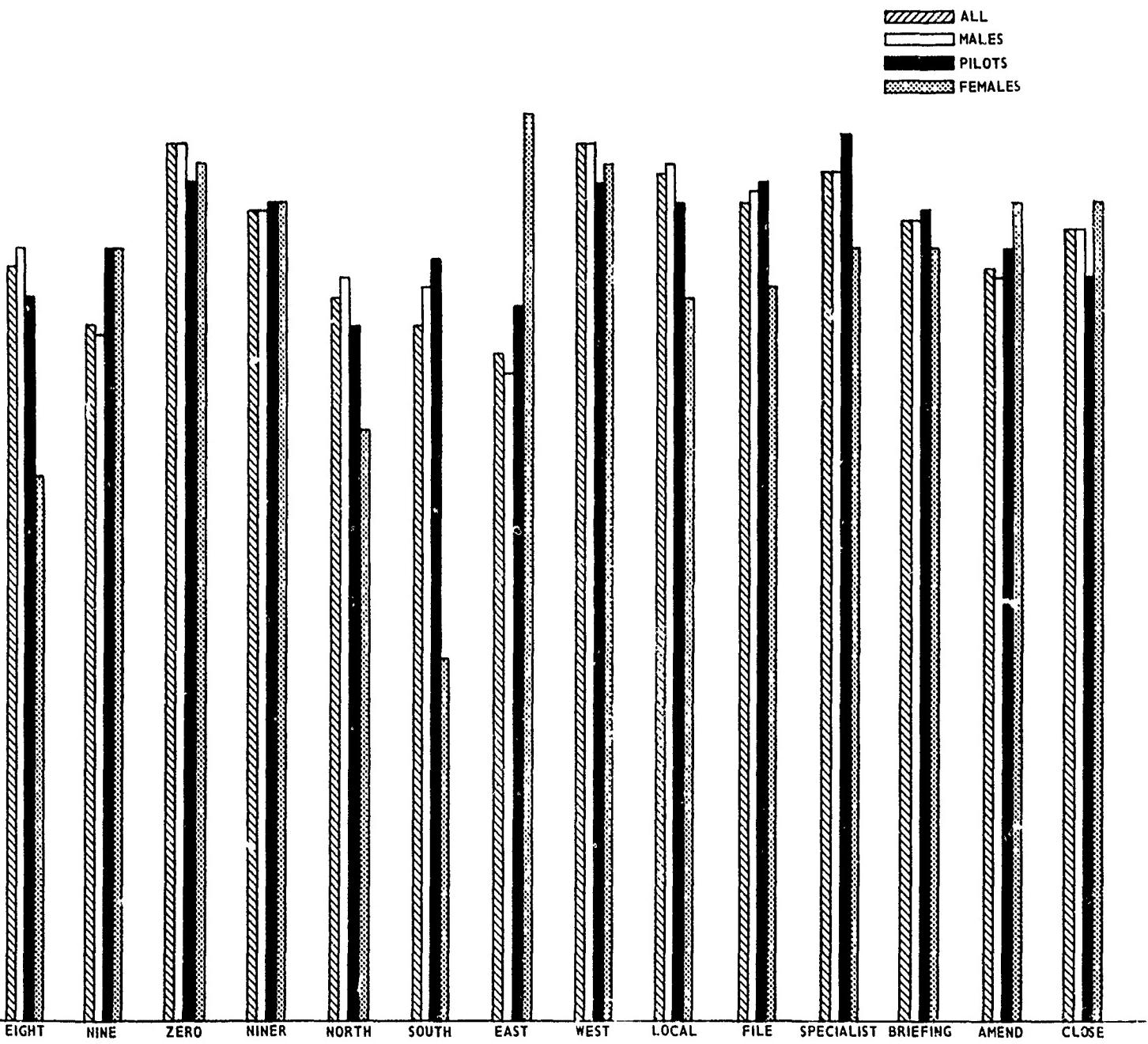


FIGURE 7. COMPARATIVE RECOGNITI



80-59-7

COMPARATIVE RECOGNITION RATE GRAP.¹

FIGURE CHOICE OF STRATEGIES

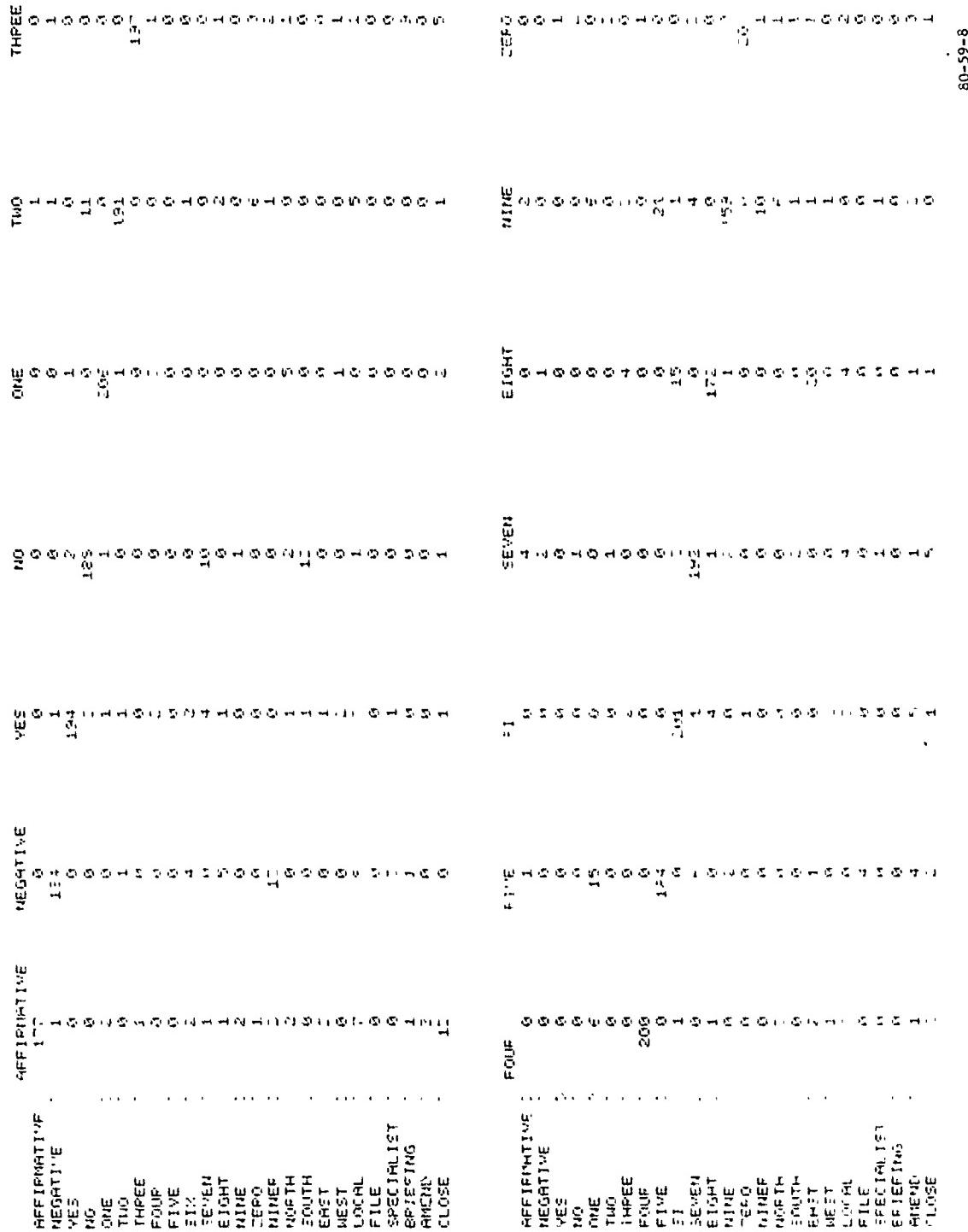


FIGURE 8. FIRST CHOICE WORD DISTRIBUTION (Sheet 1 of 2)

AFFIRMATIVE		NINEP	WEST	EAST	SOUTH	LOCAL	FILE
NEGATIVE	IS	15	1	1	6	6	6
YES	0	0	1	1	6	6	6
NO	0	0	0	1	11	6	6
ONE	0	0	1	0	0	6	6
TWO	0	0	0	0	0	6	6
THREE	1	0	0	0	0	6	6
FOUR	0	0	0	0	0	6	6
FIVE	0	0	0	0	0	6	6
SIX	0	0	0	0	0	6	6
SEVEN	0	0	0	0	0	6	6
EIGHT	0	0	0	0	0	6	6
NINE	0	0	0	0	0	6	6
CEFO	0	0	0	0	0	6	6
NINEF	0	0	0	0	0	6	6
NORTH	1	0	0	0	0	6	6
SOUTH	0	0	0	0	0	6	6
EAST	0	0	0	0	0	6	6
WEST	0	0	0	0	0	6	6
LOCAL	0	0	0	0	0	6	6
FILE	0	0	0	0	0	6	6
SPECIALIST	0	0	0	0	0	6	6
BRIEFING	0	0	0	0	0	6	6
CLOSE	0	0	0	0	0	6	6
AFFIRMATIVE		167	169	156	156	154	152
NEGATIVE		167	167	167	167	167	167
YES		167	167	167	167	167	167
NO		167	167	167	167	167	167
ONE		167	167	167	167	167	167
TWO		167	167	167	167	167	167
THREE		167	167	167	167	167	167
FOUR		167	167	167	167	167	167
FIVE		167	167	167	167	167	167
SIX		167	167	167	167	167	167
SEVEN		167	167	167	167	167	167
EIGHT		167	167	167	167	167	167
NINE		167	167	167	167	167	167
CEFO		167	167	167	167	167	167
NINEF		167	167	167	167	167	167
NORTH		167	167	167	167	167	167
SOUTH		167	167	167	167	167	167
EAST		167	167	167	167	167	167
WEST		167	167	167	167	167	167
LOCAL		167	167	167	167	167	167
FILE		167	167	167	167	167	167
SPECIALIST		167	167	167	167	167	167
BRIEFING		167	167	167	167	167	167
CLOSE		167	167	167	167	167	167
TOTALS		169	169	169	169	169	169
CLOSE		2	2	2	2	2	2
AMEND		0	0	0	0	0	0
EPIEFING		0	0	0	0	0	0
SPECIALIST		0	0	0	0	0	0
TOTALS		169	169	169	169	169	169
DATA RECORDED AND PROCESSED		169	169	169	169	169	169
BY THE FA A TECHNICAL CENTER		169	169	169	169	169	169

FIGURE 8. FIRST CHOICE WORD DISTRIBUTION (Sheet 2 of 2)

ANALYSIS.

This report is primarily concerned with that data obtained using the entire, 25-word vocabulary of the URD as valid possible choices for each interpret sequence. The reason for this is that it is estimated that to achieve direct user flight plan filing by means of utterance recognition will require, at times, a vocabulary subset consisting of control words, the numbers ZERO through NINE, including NINER and the entire phonetic alphabet with varied pronunciations. At the present time, subgroup 0, which contains all 25 vocabulary elements, most closely approximates a subgroup of the projected size. It was, however, considered valid to obtain some indication how well the URD would have performed had the test interprets been subgroup restricted. The data available from the URD at the time of testing was inadequate to derive an accurate value for the percentage of correct first choice, subgroup restricted interprets. Sufficient information was, however, available to obtain a worst case figure for the percentage of correct, subgroup restricted interprets. This figure is termed the Modified Percent Right (M%R) and may be seen in the computer print-out presented in figure 6.

The M%R is computed by adding the number of correct second choice words to the value of RIGHT in those cases where the subgroup of the incorrect first choice word was wrong. This modified version of RIGHT is then used to calculate the value of M%R. It should be stressed that the value of M%R is always less than or equal to the actual percentage of correct first choice interprets that would have been obtained if subgroup restrictions had been employed during the test. A macro flow chart of the calculation of M%R is available to the reader in appendix C.

Considering the entire data base, the overall recognition rate was increased by 5 percent by simulating subgroup restricted interprets. This results in a subgroup restricted recognition rate equal to or in excess of 90 percent.

In a standard two-pass system, the URD would have asked the caller for confirmation of his utterance approximately 33 percent of the time by initiating a WT sequence (see figure 6). Twenty-nine percent (521 cases) of these WT sequences were initiated in situations where the first choice word selected by the URD was incorrect. This means that, for the total data base, 326 incorrect words were passed to the host computer. A possibility exists that 8 of these words would be flagged by garble errors leaving 318 incorrect word codes to be passed to the host computer. This translates to 5.8 percent of all utterances transmitted to the host computer as correct will be wrong if a vocabulary of 25 words is employed. This is a best case figure.

In the case of a WT sequence, a response by the user from the AFFIRM/DENY words, subgroup 1, is required. Using the values obtained for subgroup 1 recognition, as presented in figure 6, it may be calculated that subgroup 1 has an average M%R of 93 percent. Therefore, 7 percent of all words flagged for WT confirmation will encounter an error on the confirmation word. The confirmation word is not subjected to any quality parameter tests. This percentage of inaccurately confirmed words must be added to the previous count of wrong words that escaped detection. This results in an error rate of approximately 6.5 percent for a vocabulary of 25 words, using the default quality parameters. In an ideal case, all wrong first choice words would be flagged for confirmation. This may be expressed as $WWT =$

WRONG and the AFFIRM/DENY subgroup would have a 100-percent recognition rate. Equation 1 is a means of calculating the approximate two-pass recognition rate from the information given on the computer printouts. Note the 0.93 term in equation 1 which accounts for the imperfections in AFFIRM/DENY recognition for WT sequences.

$$\frac{((WWT) \times (0.93) + RIGHT)}{RIGHT + WRONG} \times 100\% = \text{OVERALL Recognition Rate} \quad (1)$$

Assuming an ideal case in which all incorrect first choice interprets are flagged for confirmation, 100-percent recognition will not be achieved due to the imperfections in AFFIRM/DENY recognition. In actuality, the only way to ensure that all incorrect first choices are flagged for confirmation is to initiate a WT sequence for all interprets. In this case, the 93-percent AFFIRM/DENY recognition accuracy will govern the overall recognition rate. The reader should realize that the transposition of an affirm word to deny word will only result in the entry to an error handling routine. The transposition of a deny word to an affirm word in a WT sequence will result in an incorrect word code being transmitted to the host computer.

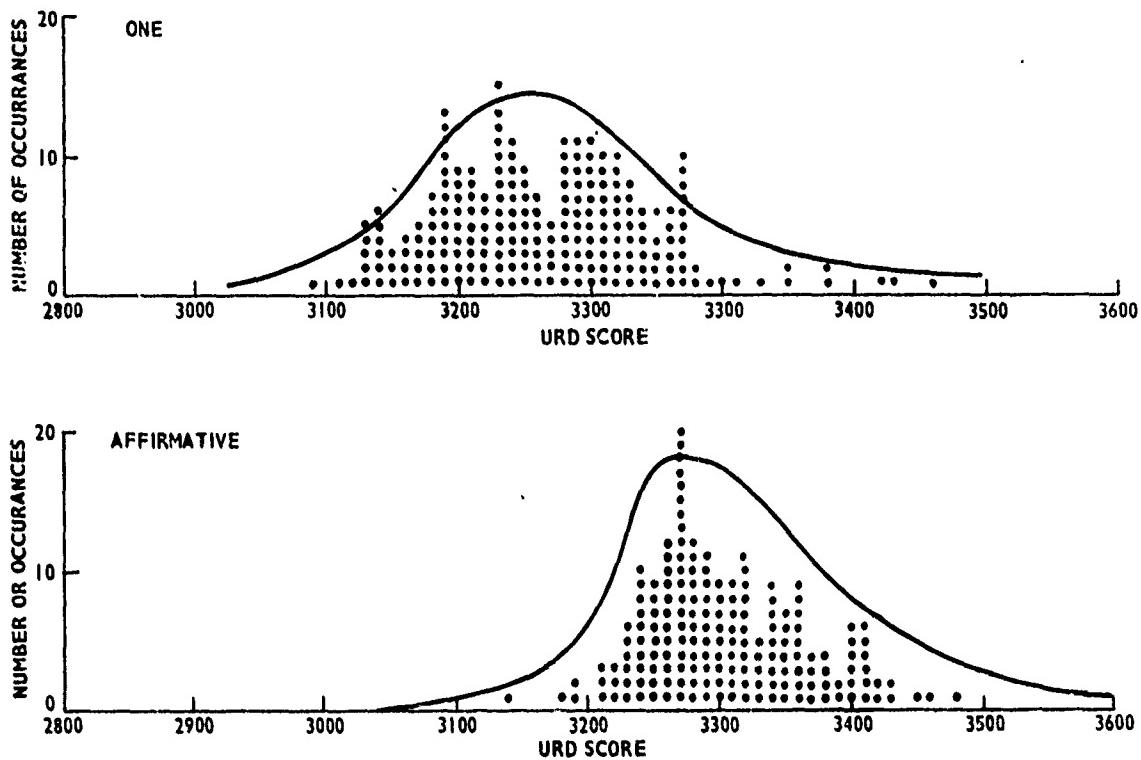
In the analysis of the score distribution, it was observed that the first choice quality scores occurred in a slightly skewed, right-handed Gaussian distribution. This trend is shown for the vocabulary elements ONE and AFFIRMATIVE in figure 9. Since the data occurs in a Gaussian fashion, the analysis of data by means of standard deviations becomes valid.

Since standard deviation analysis was shown to be valid, it was possible to

construct idealized first and second choice score distributions for each word from the data obtained in the computer analysis illustrated in figure 6. Figure 10 is an illustrated example of such an idealized graph. The method employed to construct the graph was to select the value of MEAN and DELTA for a given word and use these values as the maximum points of the first and second choice word score distribution curves. The values of STDM and STDD for the same first word were then used to construct the two idealized histograms. This was done for each word in the vocabulary. These graphs may be found in appendix B.

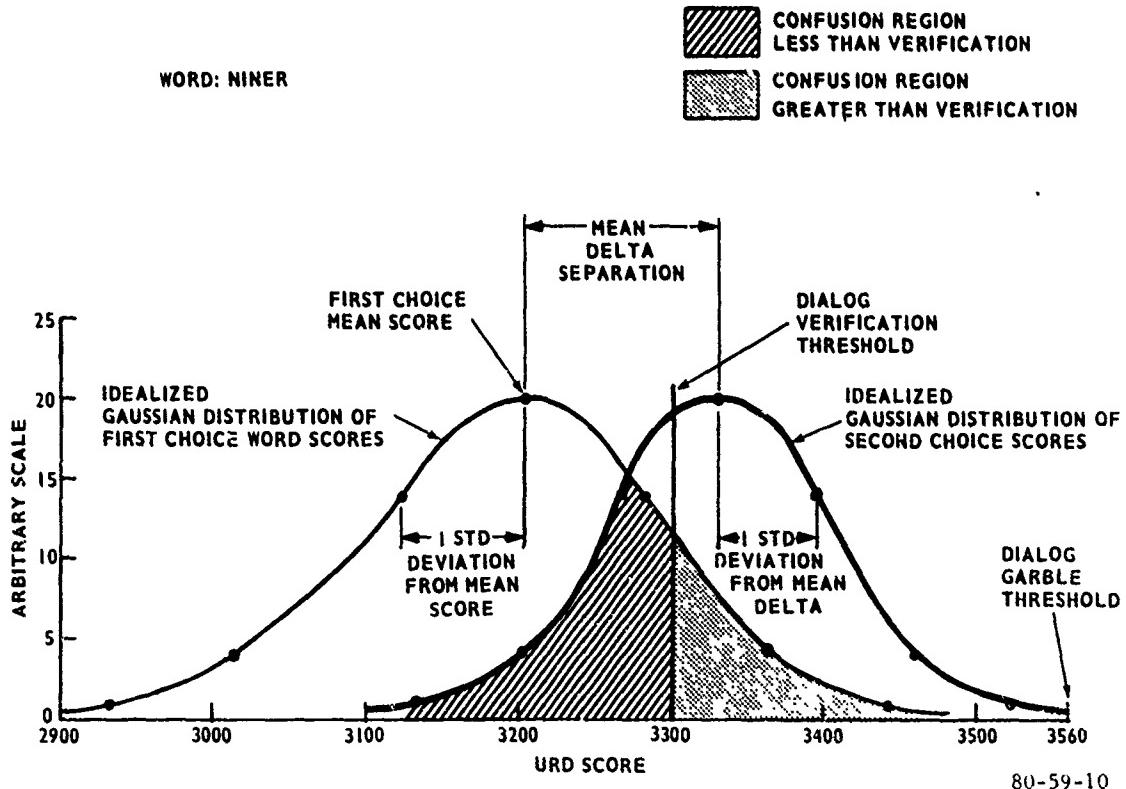
For an ideal utterance recognition device, the majority of the first choice word curve would lie to the left of the value of VERIFICATION. The separation between first and second choice words would be greater than the value of CONFUSION. Figure 11 illustrates such an idealized plot in which the above considerations are met to three standard deviations.

From figure 6 and the graphs found in appendix B, it may be noted that the mean value of the quality scores often approaches the verification thresholds for certain words. This is particularly notable for the word AFFIRMATIVE which has a mean quality score of 3,298. The default value of VERIFICATION for AFFIRMATIVE is 3,300. Referring to figure 6, it may be calculated that the word AFFIRMATIVE had a 67-percent occurrence of WT sequences. It was speculated that a modification of the default quality parameters might result in either less WT sequences, which would be more acceptable to the user from a human factors standpoint, or in an increase of WT sequences when the first choice word was incorrect. The latter would result in an increase in recognition reliability.



80-59-9

FIGURE 9. QUANTIZED SCORE DISTRIBUTION



80-59-10

FIGURE 10. IDEALIZED FIRST AND SECOND CHOICE WORD SCORE DISTRIBUTION

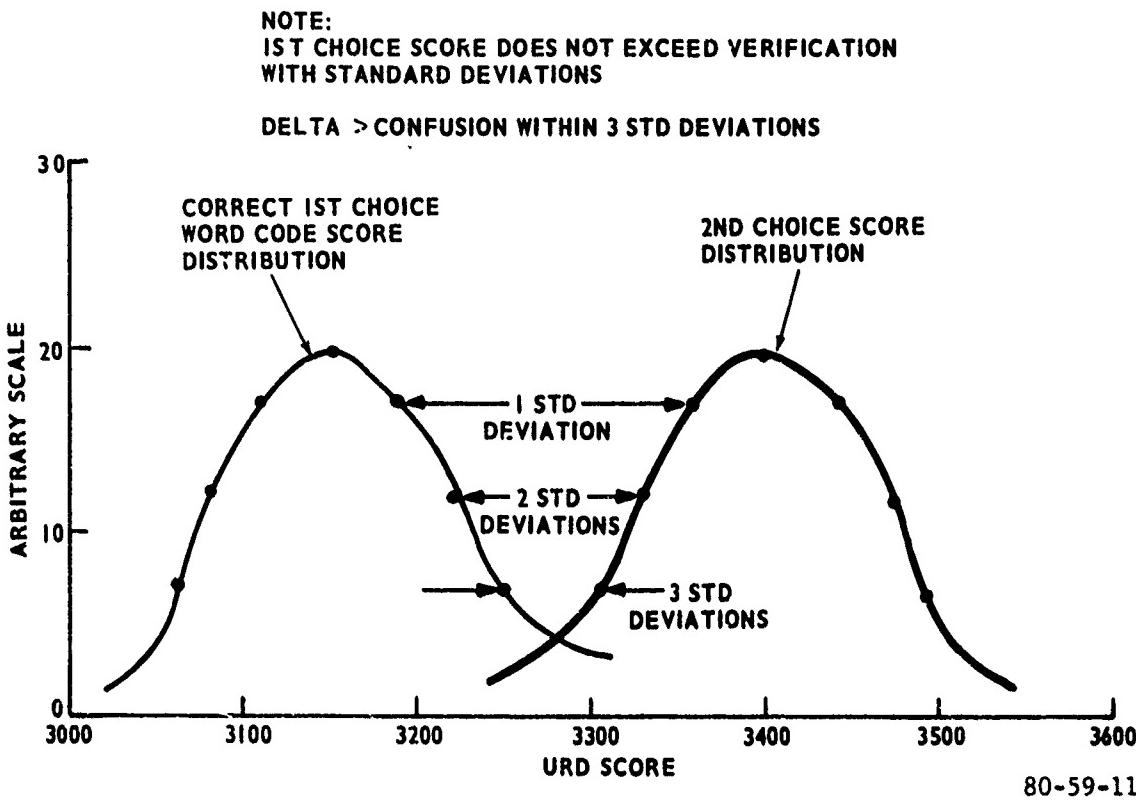


FIGURE 11. PROPOSED FIRST AND SECOND CHOICE WORD SCORE DISTRIBUTION FOR AN IDEAL UTTERANCE RECOGNITION DEVICE

An experiment was conducted; in which the value of VERIFICATION was altered to be in excess of the value of MEAN by approximately one standard deviation. The value of GARBLE was reduced by approximately one standard deviation. The value of CONFUSION was modified either up or down depending on the value of CON as determined for the particular vocabulary element. The computer analysis of the total data base, with the modified parameters, was then performed. The results showed that these changes in parameters resulted in an increase of 3 percent in the number of WT sequences that were executed by the URD. The number of wrong first choices that were flagged for confirmation only increased by approximately 1.7 percent. The

modified quality parameters as well as the corresponding results are found in appendix A. The reader should note that with the modified quality parameters, the majority of WT sequences were initiated due to confusion situations as opposed to scores in excess of the value of VERIFICATION. Also noted should be the fact that the values of RIGHT and WRONG do not change with the modification of the quality parameters.

Referring to the TOTAL column of figure 8, it may be noted that the most commonly selected first choice word is ONE. This is also true for the male and pilot subgroupings. The most common first choice word for females is LOCAL.

Analyzing the first choice distribution of figure 8, it may be noted that for many of the vocabulary elements there exists a preferred incorrect first choice word. The reader's attention is directed to the column for the word EAST in figure 8. In 42 out of 220 interprets (19 percent of the time) the word EAST was incorrectly identified as the word EIGHT. The severity of this occurrence is lessened due to the fact that the two vocabulary elements are members of different subgroups. The case of the words NINE and FIVE, (illustrated in figure 8) is far more severe owing to the fact that they may not be conveniently segregated into different subgroups.

SUMMARY OF RESULTS.

Using a vocabulary containing 25 words, the URD had a recognition rate of 85 percent. Division of the data base into subsets, according to sex and whether the subject was a pilot, showed no significant variations in terms of overall recognition from the total data base. Some individual vocabulary elements of the female subset, notably SOUTH, showed a significant variation in recognition from the predominantly male utterances.

The simulation of subgroup restricted interprets produced a recognition increase of 5 percent. This calculated increase is less than or equal to the increase that would have been achieved if subgroup restrictions had been employed during the test.

Many vocabulary elements have average quality scores extremely close to the default value of VERIFICATION resulting in an excess number of WT sequences. Modification of the quality parameters resulted in a less than 2 percent (1.7 percent) increase in the number of incorrect interprets that were flagged for confirmation. The number of WT sequences increased by 3 percent.

For many of the vocabulary elements there exists a preferred wrong first choice word. This situation may become critical when the words cannot be segregated by subgroup.

CONCLUSIONS

1. For a two-pass system, such as presently implemented using the present vocabulary with subgroup restrictions, the Utterance Recognition Device (URD) is a viable method of human/computer communications. This is true only in those cases in which a single command word is required. This was the way the URD functioned in the Mass Weather Dissemination System Exploratory Engineering Model which was the original justification for the purchase of the device.
2. Increases in the number of vocabulary elements and vocabulary subgroup size will result in a decrease in recognition reliability.
3. In an ideal case, in which all incorrect first choices are flagged for confirmation, perfect recognition will not be achieved due to the inability of the URD to perfectly recognize AFFIRM/DENY words.
4. At the present time, the URD shows a marked decrease in the recognition rate when subgroup restrictions are not employed. The filing of flight plans by utterance recognition will require a subgroup of more than 25 elements. It is projected that considering the extended subgroup size the URD's recognition rate will drop to approximately 80 percent. At this degree of recognition, all utterances will require confirmation to achieve an acceptable level of confidence. This will prove both cumbersome and time consuming to the user. Confirming all utterance will not obtain perfect recognition.

5. At the present time, the average score of an utterance when it is the correct first choice word is often sufficiently close to the value of VERIFICATION to cause an excessive number of confirmation sequences. This is unnecessary and cumbersome if the device is to be employed as a single word command input unit as in the Mass Weather System Exploratory Engineering Model.

6. There exists for many of the vocabulary elements a preferred wrong first choice word; that is, there is a given word which is most often mistaken as another vocabulary element.

RECOMMENDATIONS

1. Research in the field of untrained speech recognition over standard voice grade telephone lines should be continued. This discipline is still in its technological infancy. Continuing advancements in microelectronic technology will be reflected in future utterance recognition devices in terms of increased reliability, larger vocabulary, and lower cost per word.

2. At the present time, subgroup restricted, two-pass recognition of the existing 25-word vocabulary is a viable method of providing a general aviation user with access to an information

distribution system, such as the Mass Weather Dissemination System Exploratory Engineering Model. It is suggested that such a system be field evaluated on a limited basis to determine user acceptability of utterance recognition.

3. At the present time, it is not recommended to field evaluate a device such as the Utterance Recognition Device (URD) for direct user filing of flight plan data owing to the extreme size of the projected required vocabulary.

4. Special consideration should be given to improving the recognition rate of the AFFIRM/DENY subgroup. A reduction from four to two words in this subgroup may result in a higher degree of recognition.

5. Testing similar to that detailed in this report should be conducted with any new vocabulary or device to determine if a sufficiently high degree of recognition reliability exist.

REFERENCES

1. Shochet, E. and Lemanski, R., Automated Flight Plan Filing by Simulated Voice Recognition, FAA-RD-80-56, August 1980.

APPENDIX A

COMPUTER ANALYSES (INCLUDING GLOSSARY OF TERMS)

This appendix contains a complete set of computer analyses for the four major data base subdivisions — the total data base, all males, all pilots, and all females. The data contained in this appendix has been briefly summarized in table 1 and figure 7 which appear in the body of the report. A Glossary of Terms is included at the end of this appendix.

Figure A-1A analyses the total data base and includes an analysis of the same data subset using the modified quality parameters. The computer analysis using modified quality parameters is given only for the total data base. The reader should note that there is no difference in the terms which represent the absolute functions of the URD (RIGHT, WRONG, %R, and MZR). Only those terms which reflect confirmation are changed (WT, PR, WWT, and CON).

Figure A-1B is a computer-generated bargraph of the relative recognition rate of each word. The figure shows the percentage of correct first choices. (Figure 7 in the text is a summary of these graphs for all four data subsets.) Figure A-1C is a similar graph representing recognition rates based on the simulated subgroup restrictions. This figure shows the modified percent right (MZR).

Figure A-1D is a comparison of the relative quality scores for each vocabulary element based on the value of MEAN. To allow relatively compact graphing, this has been presented as one-tenth of the value of MEAN greater than 3,000 ((MEAN-3,000)/10). The reader should note the tendency of some vocabulary elements to approach the default value of VERIFICATION, 3,300.

Figure A-1E is the first choice word distribution. This indicates how many times each vocabulary element was selected as the first choice word in a given situation. Column headers indicate the expected word.

The computer analyses for males are found in figure A-2; pilots, figure A-3; and females, A-4.

GLOSSARY OF TERMS

AMPR — This term represents the arithmetic average of the amplitudes of the correct first choice interprets.

AMPW — This term represents the arithmetic average of the amplitudes of incorrect first choice utterances.

CON — This term represents the number of times that the URD would have asked the speaker "What was that —?" due to a confusion situation that was not overridden by a verification or GARBLE situation using a given set of parameters.

CONFUSION — If the difference between the scores of the first and second choice words is less than or equal to this Dialog system parameter, the URD will ask the speaker "Was that —?"

DELTA — This term represents the arithmetic average of the difference between the first and second choice scores when the first choice word is correct.

GARBLE — If the first choice score is greater than or equal to this Dialog system parameter, the URD will ask the speaker to please repeat the previously stated word. If this parameter is exceeded on the second attempt, the URD will ask the speaker "Was that —?" Where — is the first choice word. Garble conditions have priority over verification conditions.

MEAN — This term represents the arithmetic average of the correct first choice scores.

M%R — This term represents an approximation of the correct first choice percentage if the URD test had been conducted using subgroups. A modified right score is generated by adding the number of times that the second choice word was correct logically and with an incorrect first choice subgroup to **RIGHT**. **M%R** must always be considered as being less than or equal to the actual percentage correct that would have been obtained by using subgroups.

%R — This term represents the percentage of correct first choices.

RIGHT — This number represents the number of times that the URD's first choice word was correct. In the case of a please repeat situation, the results of both the first and second interprets are considered to be valid data.

SEC — This term represents the total number of times that the URD's second choice was the actual word said by the subject.

SGW — This term represents the total number of times that the first choice was a member of a subgroup different than that of the correct word.

STDD — This term represents the standard deviation from **DELTA** for all correct first choice cases.

STD_M — This term represents the standard deviation from **MEAN** of the scores of the correct first choice interprets.

STD_R — This term represents the standard deviation of the correct first choice interpret amplitudes from **AMPR**.

STDW — This term represents the standard deviation of the incorrect first choice utterances from **AMPW**.

TO — This term represents the total number of times that a time-out condition occurred during an interpret. A time-out situation occurs when the URD does not receive an audio input of sufficient amplitude and duration within 5 seconds after the beginning of an interpret. The URD handles this situation by asking the subject to repeat the word previously said. Two consecutive time-outs cause the test sequence to be aborted.

VERIFICATION — If the first choice score is greater than or equal to this Dialog system parameter, the URD will ask the speaker "Was that —?" Verification conditions have priority over confusion conditions.

WRONG — This number represents the total number of times that the URD's first choice word was incorrect. The "please repeat" situation is the same as for **RIGHT**.

WT — This term represents the number of times that the URD would have asked the speaker "Was that —?" using a given set of parameters.

WWT — This term represents the number of times that the URD would have asked "Was that —?" when the first choice word was incorrect.

ALL	UPD	GAPBLE	VERIFICATION	CONFUSION
WORD				
AFFIRMATIVE		2560	1309	59
NEGATIVE		2560	1269	59
YES		2560	1265	59
NO		2560	1269	59
ONE		2560	1269	59
TWO		2560	1269	59
THREE		2560	1268	59
FOUR		2560	1269	59
FIVE		2560	1269	59
SIX		2560	1269	59
SEVEN		2560	1269	59
EIGHT		2560	1269	59
NINE		2560	1268	59
ZERO		2560	1269	59
NINEP		2560	1269	59
NOPIN		2560	1269	59
SOUTH		2560	1269	59
EAST		2560	1269	59
WEST		2560	1269	59
LOCAL		3560	2300	59
FILE		3560	2300	59
SPECIALIST		3560	2300	59
BRIEFING		3560	2300	59
AMEND		3560	2300	59
CLOSE		3560	2300	59
WORD		RIGHT	WRONG	WT
AFFIRMATIVE		177	43	148
NEGATIVE		184	25	162
YES		194	31	43
NO		188	31	35
ONE		206	13	25
TWO		194	29	74
THREE		197	22	40
FOUR		200	19	34
FIVE		184	35	57
SIX		201	19	58
SEVEN		192	27	56
EIGHT		172	47	110
NINE		159	60	81
ZERO		202	17	27
NINEP		187	21	64
NOPIN		167	52	162
SOUTH		168	59	85
EAST		155	65	112
WEST		203	17	79
LOCAL		194	24	6
FILE		189	31	64
SPECIALIST		194	25	84
BRIEFING		185	35	75
AMEND		173	46	94
CLOSE		182	37	117
TOTALS		4635	847	1865
TOTAL INTERRUPTS				>5483
PERCENT CORRECT				85
PERCENT WAS THAT				12

FIGURE A-1A. TOTAL DATA BASE — COMPUTER ANALYSIS (Sheet 1 of 2)

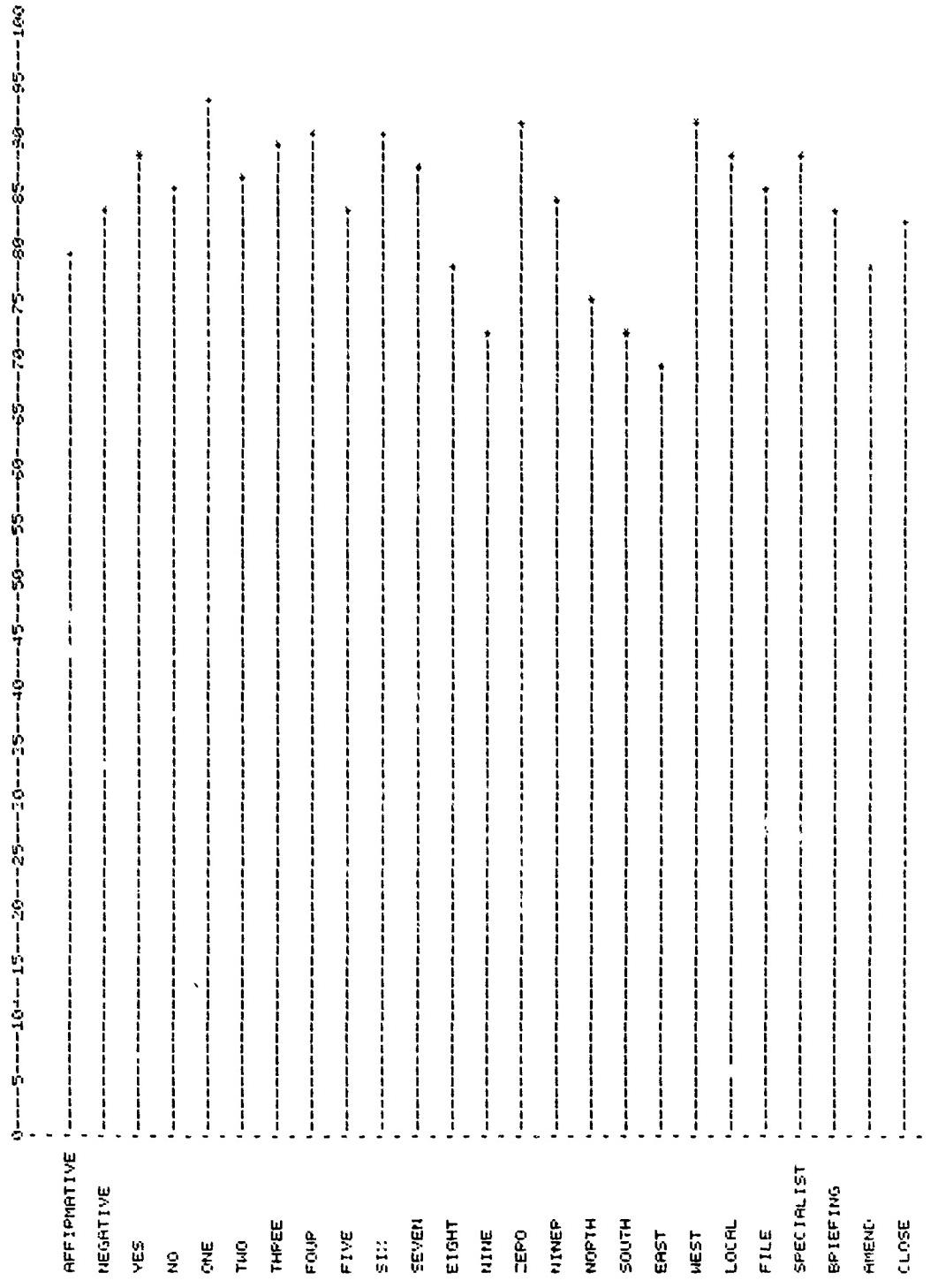
80-59-A-1A

ALL • UPD	WORD	VERIFICATION	CONFUSION
AFFIRMATIVE	> 25000	23555	65
NEGATIVE	> 24999	23499	55
YES	> 2410	22445	60
NO	> 2400	22440	60
ONE	> 2420	22600	30
TWO	> 2420	22520	65
THREE	> 2450	22448	75
FOUR	> 2420	2276	45
FIVE	> 2370	2228	50
SIX	> 2500	2146	50
SEVEN	> 2440	2268	70
EIGHT	> 2510	2328	60
NINE	> 2390	2268	60
TEN	> 2450	22440	70
NINEP	> 2470	2288	80
NORTH	> 2440	2288	70
SOUTH	> 2450	2360	75
EAST	> 2500	2248	70
WEST	> 2490	2268	30
LOCAL	> 2520	2248	75
FILE	> 2410	2268	45
SPECIALIST	> 2500	2350	55
BRIEFING	> 2540	2348	80
AMEND	> 2500	2346	45
CLOSE	> 2500	2330	25
WORD	WRONG	WIT	CON
AFFIRMATIVE	> 177	171	PP
NEGATIVE	> 184	184	TO
YES	> 194	194	SEC
NO	> 188	21	PP
ONE	> 206	51	6
TWO	> 191	43	6
THREE	> 197	29	5
FOUR	> 200	19	4
FIVE	> 184	35	2
SIX	> 201	49	2
SEVEN	> 192	75	2
EIGHT	> 172	47	2
NINE	> 159	60	1
TEN	> 202	47	1
NINEP	> 187	72	0
NORTH	> 167	52	0
SOUTH	> 169	59	0
EAST	> 155	65	1
WEST	> 203	17	1
LOCAL	> 154	65	0
FILE	> 189	59	1
SPECIALIST	> 194	25	1
BRIEFING	> 155	75	1
AMEND	> 172	46	1
CLOSE	> 182	37	1
TOTALS	> 45326	847	1939
TOTAL INTERPRETS	> 5483		
PERCENT CORRECT	> 85		
PERCENT WAS THAT	> 36		

FIGURE A-1A. TOTAL DATA BASE — COMPUTER ANALYSIS (Sheet 2 of 2)

VERIFIED PARAMETERS

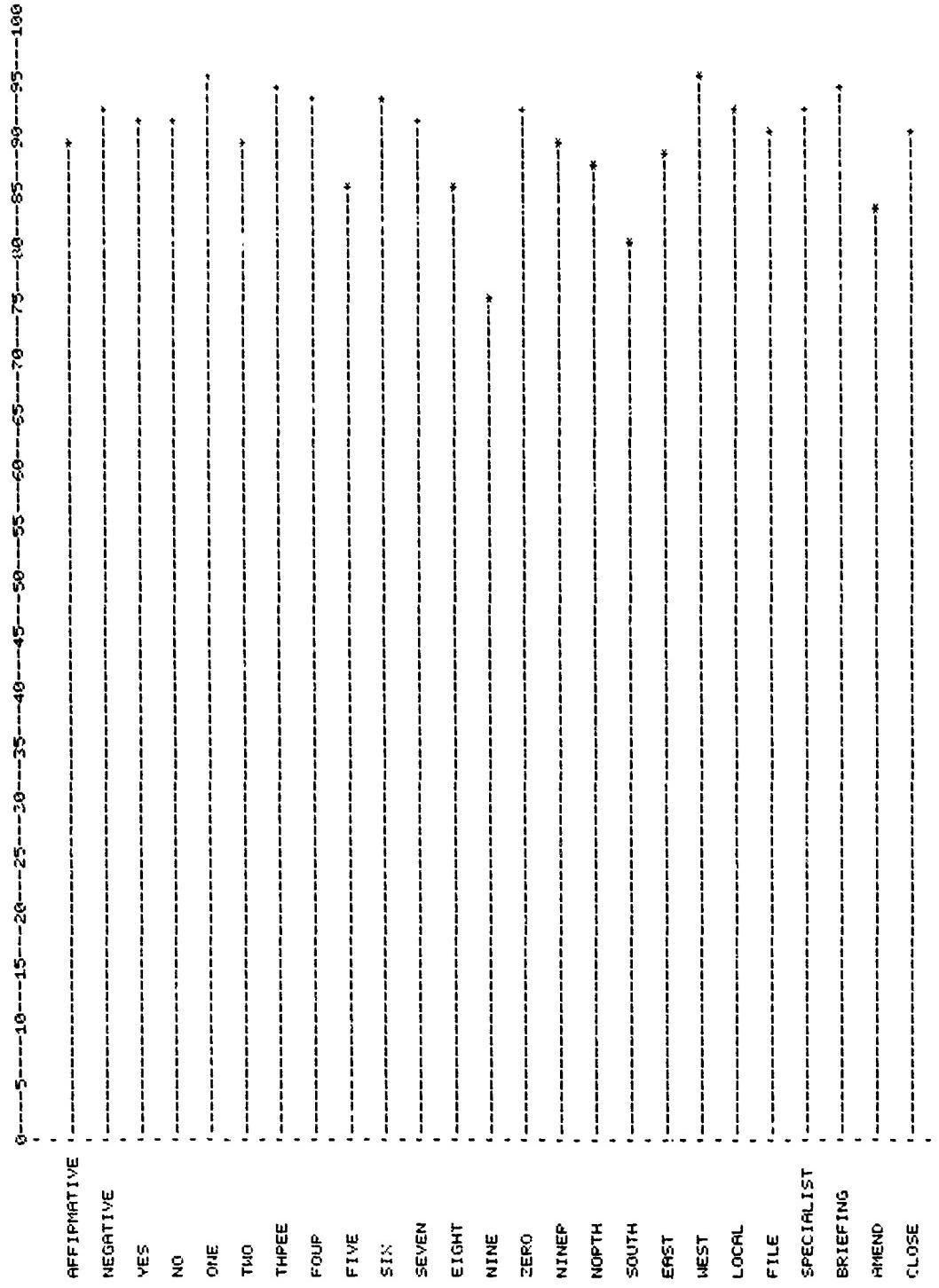
80-59-A-1A



• PERCENTAGE OF CORRECT FIRST CHOICES

80-59-A-1B

FIGURE A-1B. TOTAL DATA BASE — RELATIVE RECOGNITION RATE



MODIFIED PERCENT CORRECT
80-59-A-1C

FIGURE A-1C. TOTAL DATA BASE — SIMULATED SUBGROUP RESTRICTIONS

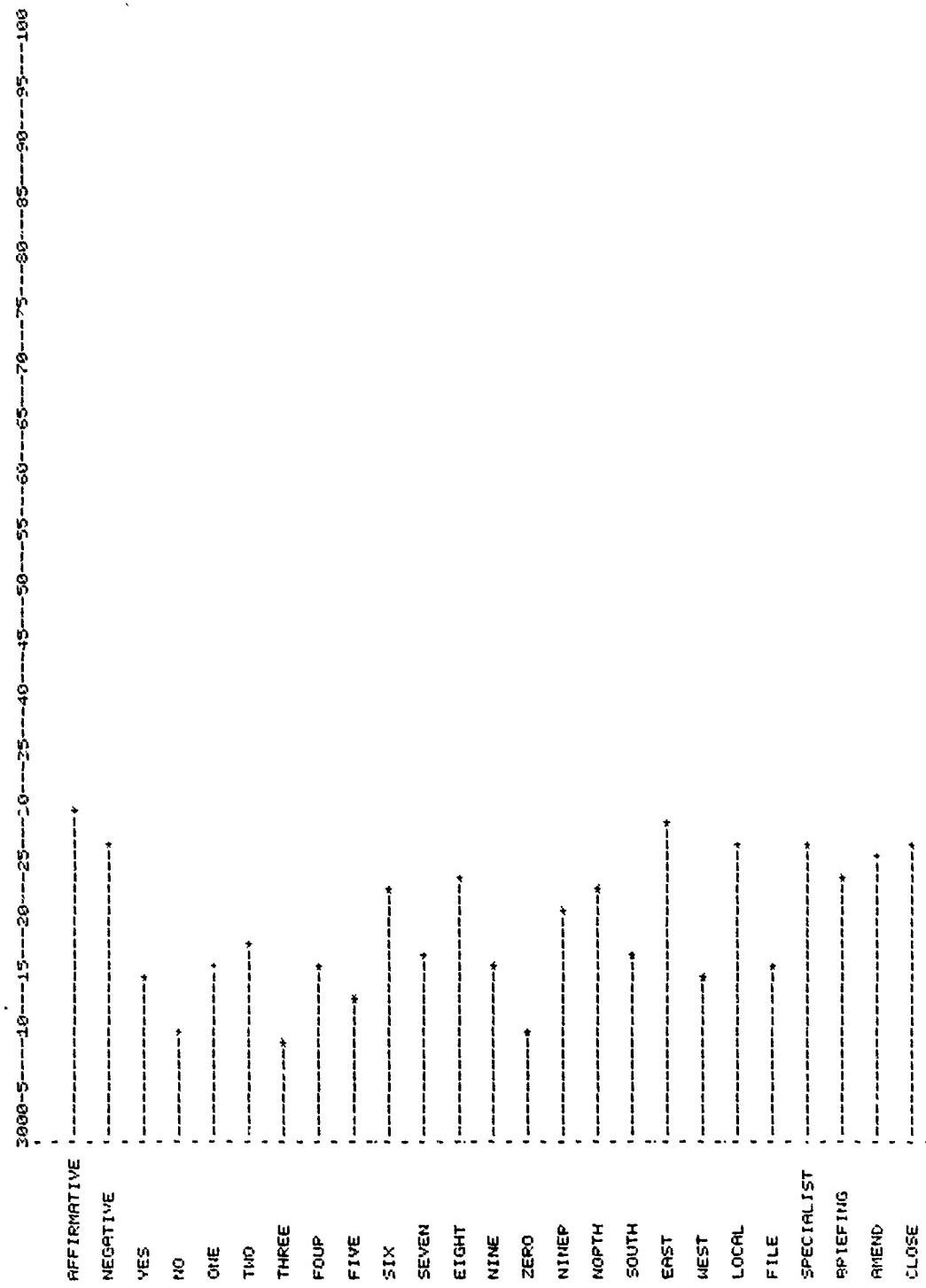


FIGURE A-1D. TOTAL DATA BASE — COMPARISON OF RELATIVE QUALITY SCORES

1718 MEAN SCORE OVER 2000 80-59-A-1D

FIRST CHOICE DISTRIBUTION

AFFIRMATIVE	>	177	NEGATIVE	>	0	NO	>	0	ONE	>	0	TWO	>	1	THREE	>	1
NEGATIVE	>	1	YES	>	184	NO	>	1	NO	>	1	NO	>	1	NO	>	1
NO	>	0	ONE	>	0	ONE	>	1	NO	>	2	NO	>	1	NO	>	1
ONE	>	2	TWO	>	0	NO	>	1	NO	>	1	NO	>	1	NO	>	1
TWO	>	0	THREE	>	1	NO	>	1	NO	>	1	NO	>	1	NO	>	1
THREE	>	0	FOUR	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
FOUR	>	0	FIVE	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
FIVE	>	0	SIX	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
SIX	>	0	SEVEN	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
SEVEN	>	0	EIGHT	>	1	NO	>	1	NO	>	1	NO	>	1	NO	>	1
EIGHT	>	0	NINE	>	1	NO	>	1	NO	>	1	NO	>	1	NO	>	1
NINE	>	0	ZERO	>	1	NO	>	1	NO	>	1	NO	>	1	NO	>	1
ZERO	>	0	NINEP	>	1	NO	>	1	NO	>	1	NO	>	1	NO	>	1
NINEP	>	0	NORTH	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
NORTH	>	0	SOUTH	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
SOUTH	>	0	EAST	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
EAST	>	0	WEST	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
WEST	>	0	LOCAL	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
LOCAL	>	0	FILE	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
FILE	>	0	SPECIALIST	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
SPECIALIST	>	0	BRIEFING	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
BRIEFING	>	0	AMEND	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
AMEND	>	0	CLOSE	>	2	NO	>	2	NO	>	2	NO	>	2	NO	>	2
CLOSE	>	0		>	0		>	0		>	0		>	0		>	0

AFFIRMATIVE	>	177	NEGATIVE	>	0	NO	>	0	ONE	>	0	NO	>	0	ZERO	>	0
NEGATIVE	>	1	YES	>	184	NO	>	1	NO	>	2	NO	>	2	ZERO	>	0
NO	>	0	ONE	>	0	NO	>	1	NO	>	1	NO	>	1	ZERO	>	0
ONE	>	2	TWO	>	0	NO	>	1	NO	>	1	NO	>	1	ZERO	>	0
TWO	>	0	THREE	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
THREE	>	0	FOUR	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
FOUR	>	0	FIVE	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
FIVE	>	0	SIX	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
SIX	>	0	SEVEN	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
SEVEN	>	0	EIGHT	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
EIGHT	>	0	NINE	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
NINE	>	0	ZERO	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
ZERO	>	0	NINEP	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
NINEP	>	0	NORTH	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
NORTH	>	0	SOUTH	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
SOUTH	>	0	EAST	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
EAST	>	0	WEST	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
WEST	>	0	LOCAL	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
LOCAL	>	0	FILE	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
FILE	>	0	SPECIALIST	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
SPECIALIST	>	0	BRIEFING	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
BRIEFING	>	0	AMEND	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
AMEND	>	0	CLOSE	>	2	NO	>	2	NO	>	2	NO	>	2	ZERO	>	0
CLOSE	>	0		>	0		>	0		>	0		>	0		>	0

FIGURE A-1E. TOTAL DATA BASE — FIRST CHOICE WORD DISTRIBUTION (Sheet 1 of 2)

80-59-A-1E

**DATA RECORDED AND PROCESSED
BY THE FAA TECHNICAL CENTER**

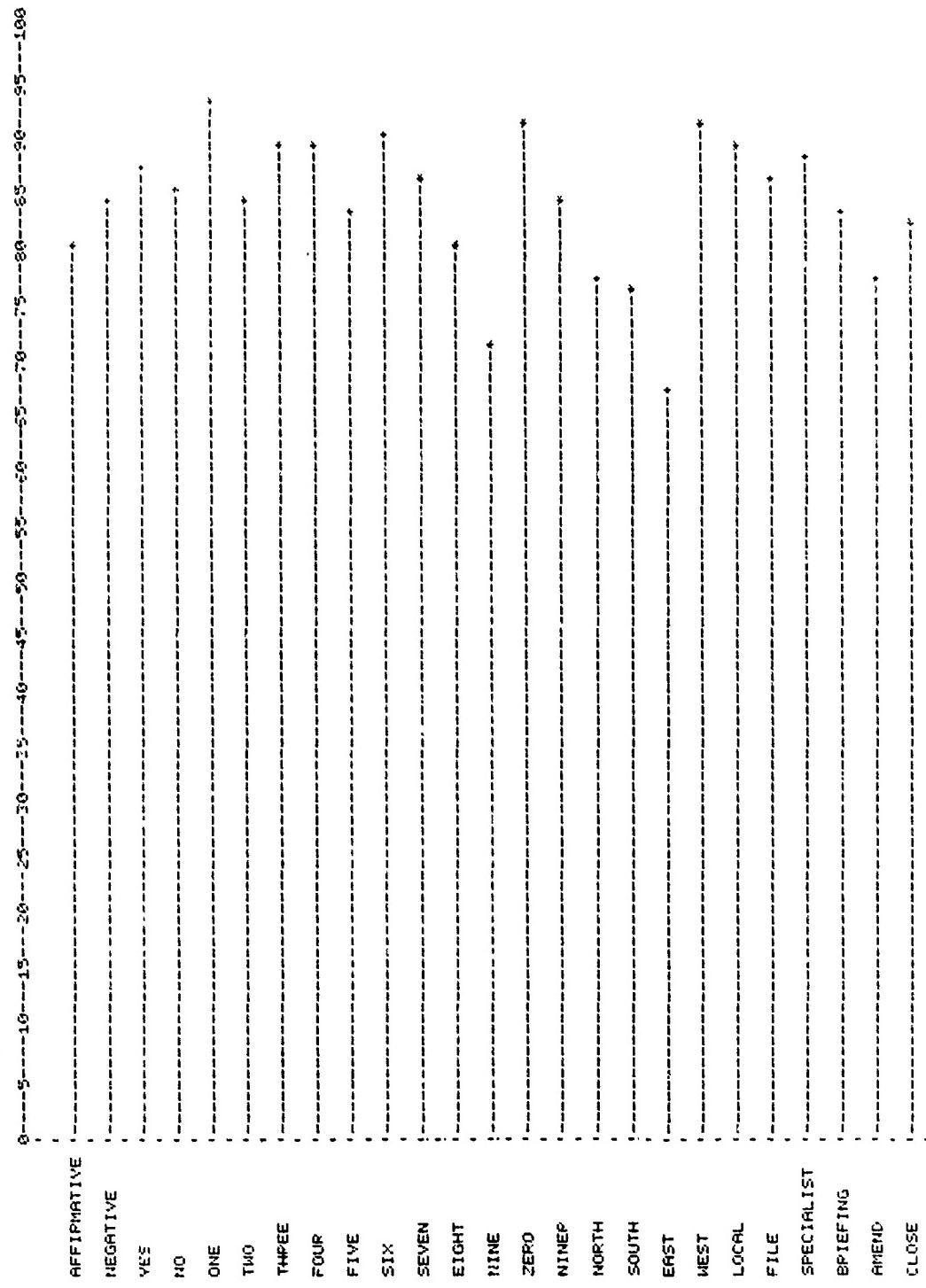
80 - 29 - A - 1E

FIGURE A-1E. TOTAL DATA BASE — FIRST CHOICE WORD DISTRIBUTION (Sheet 2 of 2)

DATA RECORDED AND PROCESSED
BY THE FAA TECHNICAL CENTER

MALES • URC	WORD	CABLE	VERIFICATION	CONFUSION	
AFFIRMATIVE	2560	2260	50		
NEGATIVE	2560	2260	50		
YES	3560	3260	50		
NO	3560	3260	50		
ONE	3560	3260	50		
TWO	2560	2260	50		
THREE	2560	2260	50		
FOUR	3560	3260	50		
FIVE	2560	2260	50		
SIX	3560	3260	50		
SEVEN	2560	2260	50		
EIGHT	2560	2260	50		
NINE	2560	2260	50		
ZERO	2560	2260	50		
HINER	3560	3260	50		
NORTH	2560	3260	50		
SOUTH	2560	3260	50		
EAST	2560	3260	50		
WEST	2560	3260	50		
LOCAL	3560	3260	50		
FILE	2560	2260	50		
SPECIALIST	2560	2260	50		
BRIEFING	3560	3260	50		
AMEND	2560	2260	50		
CLOSE	3560	3260	50		
WORD	RIGHT	WRONG	WT	TO SEC	
AFFIRMATIVE	162	37	135	1	
NEGATIVE	163	23	67	0	
YES	175	23	19	12	
NO	170	28	20	14	
ONE	186	12	22	0	
TWO	170	29	50	1	
THREE	179	19	32	0	
FOUR	179	19	29	0	
FIVE	166	32	51	0	
SIX	182	17	54	1	
SEVEN	172	26	52	0	
EIGHT	160	38	97	0	
NINE	142	96	75	0	
ZERO	183	15	24	0	
HINER	169	19	57	1	
NORTH	154	44	88	0	
SOUTH	152	46	74	0	
EAST	135	64	126	1	
WEST	184	15	26	1	
LOCAL	178	29	83	0	
FILE	172	26	50	0	
SPECIALIST	177	21	74	0	
BRIEFING	168	34	65	1	
AMEND	155	43	87	0	
CLOSE	164	24	104	0	
TOTALS	4280	754	1613	7	
TOTAL INTERPRETS	74957				
PERCENT CORRECT	85				
PERCENT WAS THAT	37				

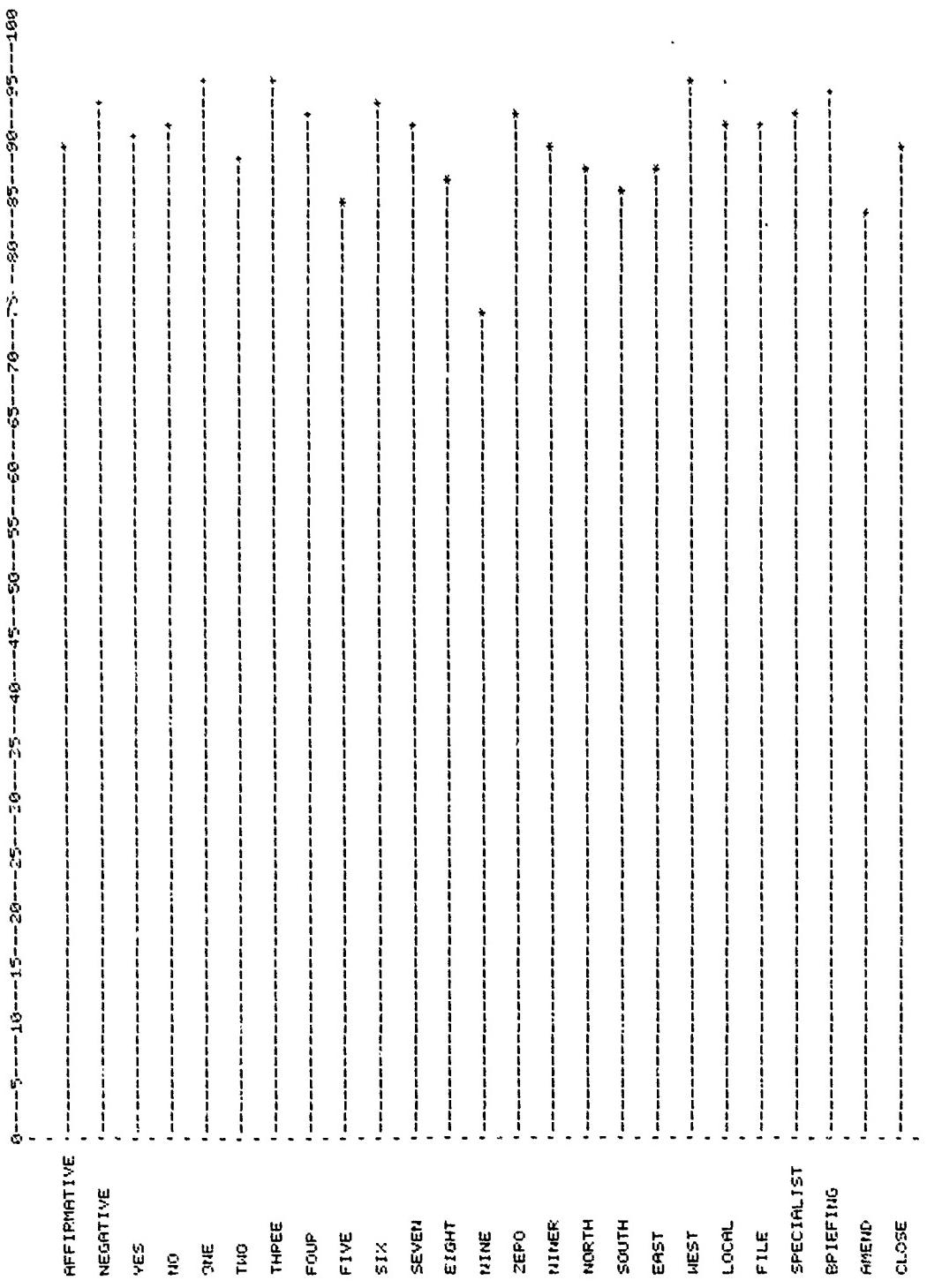
FIGURE A-2A. MALES — COMPUTER ANALYSIS



A-11

PERCENTAGE OF CORRECT FIRST CHOICES
80-59-A-2B

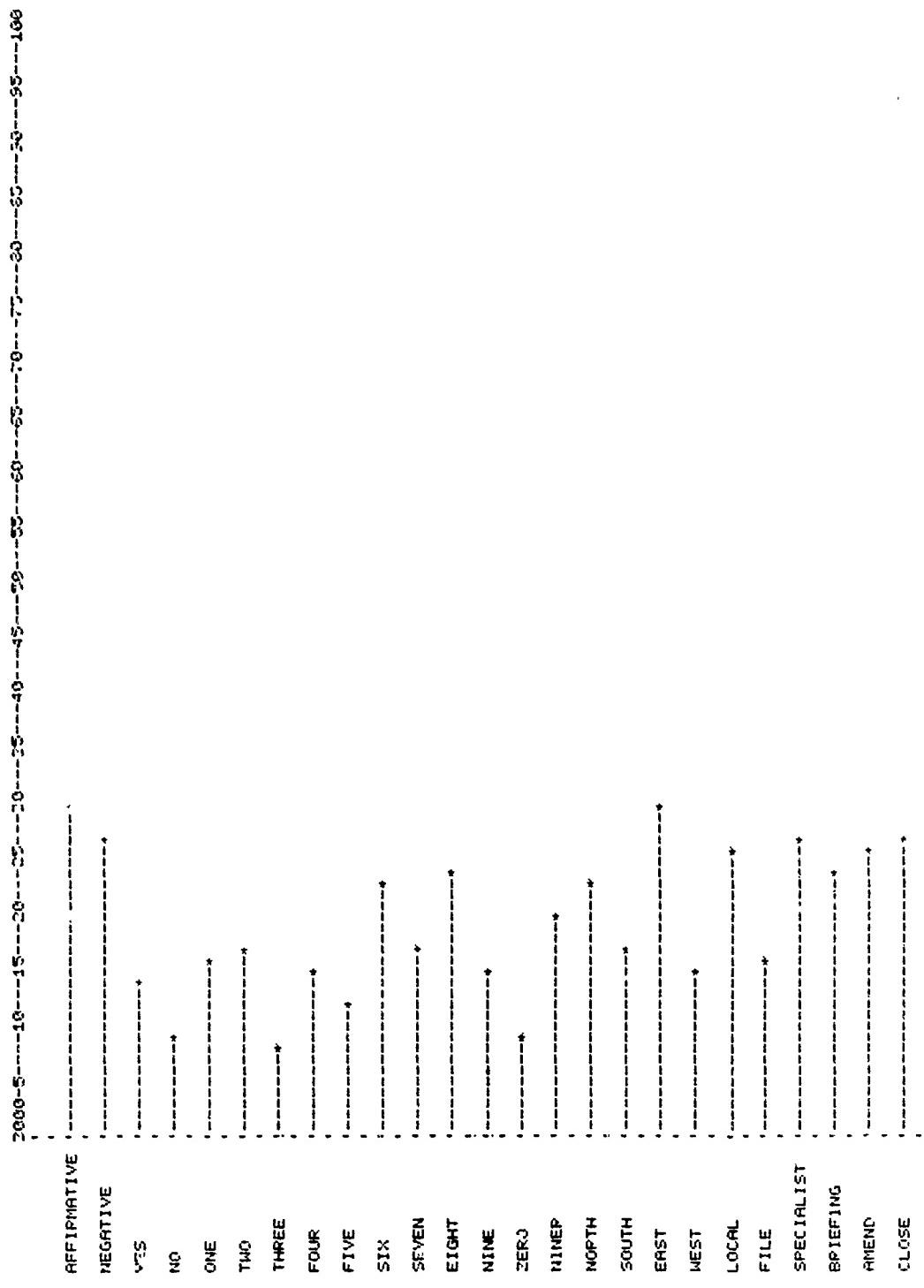
FIGURE A-2B. MALES — RELATIVE RECOGNITION RATE



MODIFIED PERCENT CORRECT

80-59-A-2C

FIGURE A-2C. MALES — SIMULATED SUBGROUP RESTRICTIONS



A-13

1/10 KERN SCORE OVER 3000

80-59-A-2D

FIGURE A-2D. MATES - COMPARISON OF RELATIVE QUALITY SCORES

FIRST CHOICE DISTRIBUTION

	AFFIRMATIVE	NEGATIVE	NO	ONE	TWO	THREE
AFFIRMATIVE	162	169	6	6	1	1
NEGATIVE	16	16	1	1	1	1
YES	10	10	1	1	1	1
NO	10	10	1	1	1	1
ONE	10	10	1	1	1	1
TWO	10	10	1	1	1	1
THREE	10	10	1	1	1	1
FOUR	10	10	1	1	1	1
FIVE	10	10	1	1	1	1
SIX	10	10	1	1	1	1
SEVEN	10	10	1	1	1	1
EIGHT	10	10	1	1	1	1
NINE	10	10	1	1	1	1
ZERO	10	10	1	1	1	1
NINEP	10	10	1	1	1	1
NORTH	10	10	1	1	1	1
SOUTH	10	10	1	1	1	1
EAST	10	10	1	1	1	1
WEST	10	10	1	1	1	1
LOCAL	10	10	1	1	1	1
FILE	10	10	1	1	1	1
SPECIALIST	10	10	1	1	1	1
BRIEFING	10	10	1	1	1	1
AMEND	10	10	1	1	1	1
CLOSE	10	10	1	1	1	1
	12	12	12	12	12	12
	11	11	11	11	11	11
	10	10	10	10	10	10
	9	9	9	9	9	9
	8	8	8	8	8	8
	7	7	7	7	7	7
	6	6	6	6	6	6
	5	5	5	5	5	5
	4	4	4	4	4	4
	3	3	3	3	3	3
	2	2	2	2	2	2
	1	1	1	1	1	1
	0	0	0	0	0	0

80-59-A-2E

FIGURE A-2E. MALES — FIRST CHOICE WORD DISTRIBUTION (Sheet 1 of 2)

FIGURE A-2E. MALES -- FIRST CHOICE WORD DISTRIBUTION (Sheet 2 of 2)

PILOTS - UPD

CONFUSION

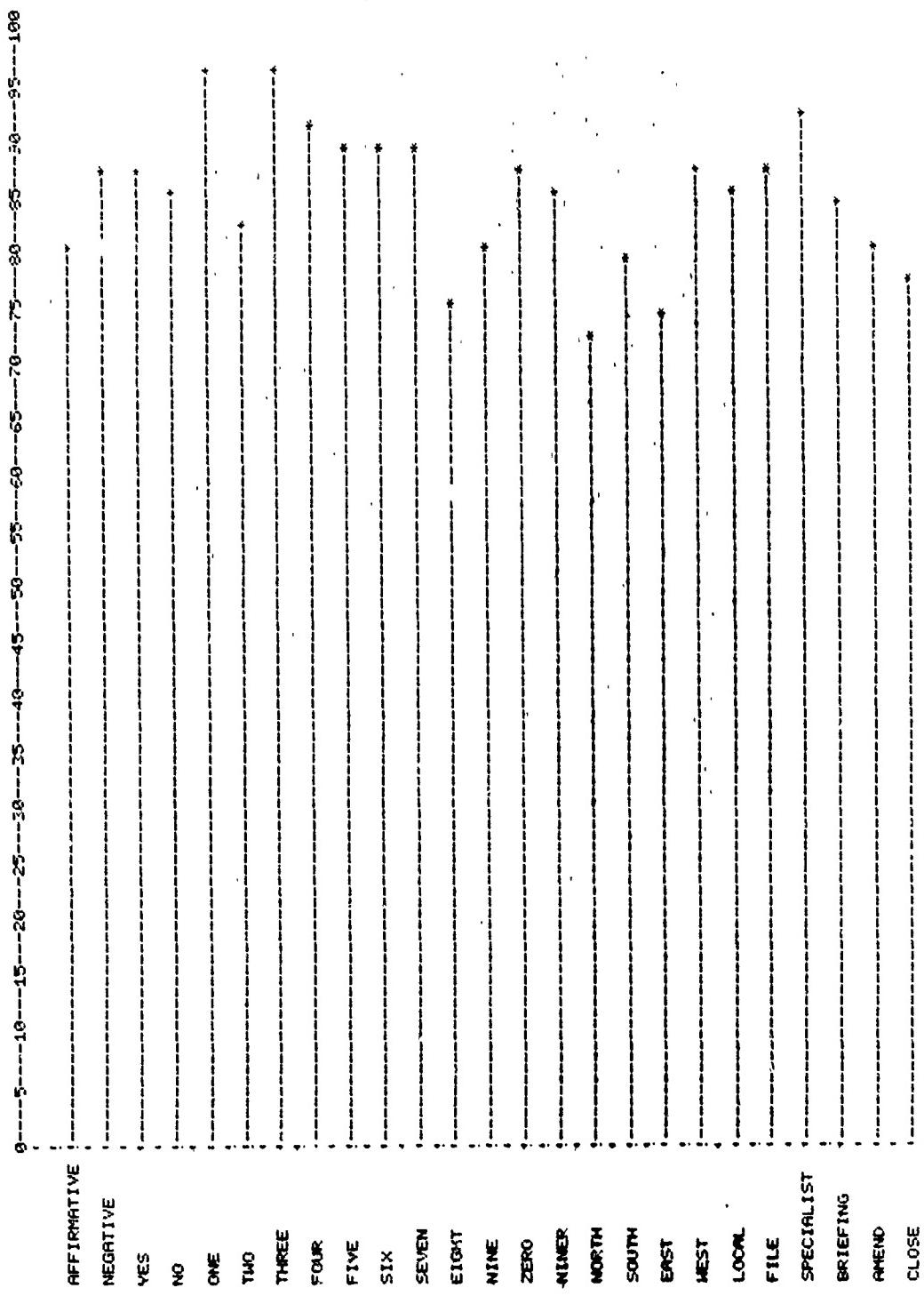
DATA RECORDED AND PROCESSED
BY THE FAA TECHNICAL CENTER

WORD	GAPELLE	VERIFICATION	CONFUSION
AFFIRMATIVE	3569	2188	56
NEGATIVE	1561	2106	59
YES	1561	2106	59
NO	2566	2160	59
ONE	1560	2160	59
TWO	1560	2160	59
THREE	2566	2160	59
FOUR	1560	2160	59
FIVE	1560	2160	59
SIX	1560	2160	59
SEVEN	1560	2160	59
EIGHT	2566	2160	59
NINE	1560	2160	59
ZERO	1560	2160	59
NINER	2566	2160	59
NORTH	1560	2160	59
SOUTH	2566	2160	59
EAST	1560	2160	59
WEST	2566	2160	59
LOCAL	1560	2160	59
FILE	1560	2160	59
SPECIALIST	3560	2160	59
BRIEFING	2566	2160	59
ARMED	1560	2160	59
CLOSE	2566	2160	59
WORD	RIGHT	WPOINT	WT
AFFIRMATIVE	48	11	41
NEGATIVE	52	7	2
YES	52	7	11
NO	54	6	11
ONE	57	2	4
TWO	49	10	19
THREE	57	5	15
FOUR	54	16	12
FIVE	53	6	12
SIX	52	6	15
SEVEN	52	6	14
EIGHT	45	14	21
NINE	49	11	19
ZERO	52	7	7
NINER	51	8	19
NORTH	43	16	20
SOUTH	47	12	17
EAST	44	15	27
WEST	52	7	16
LOCAL	51	8	25
FILE	52	7	12
SPECIALIST	55	4	17
BRIEFING	51	9	18
ARMED	48	11	25
CLOSE	46	13	24
TOTALS	>1264	212	462
TOTAL INTERPRETS	>1476		
PERCENT CORRECT	> 86		
PERCENT WAS THAT	> 71		

TOTAL INTERPRETS
PERCENT CORRECT
PERCENT WAS THAT

FIGURE A-3A.

PILOTS — COMPUTER ANALYSIS



PERCENTAGE OF CORRECT FIRST CHOICES

80-59-A-3B

FIGURE A-3B. PILOTS — RELATIVE RECOGNITION RATE

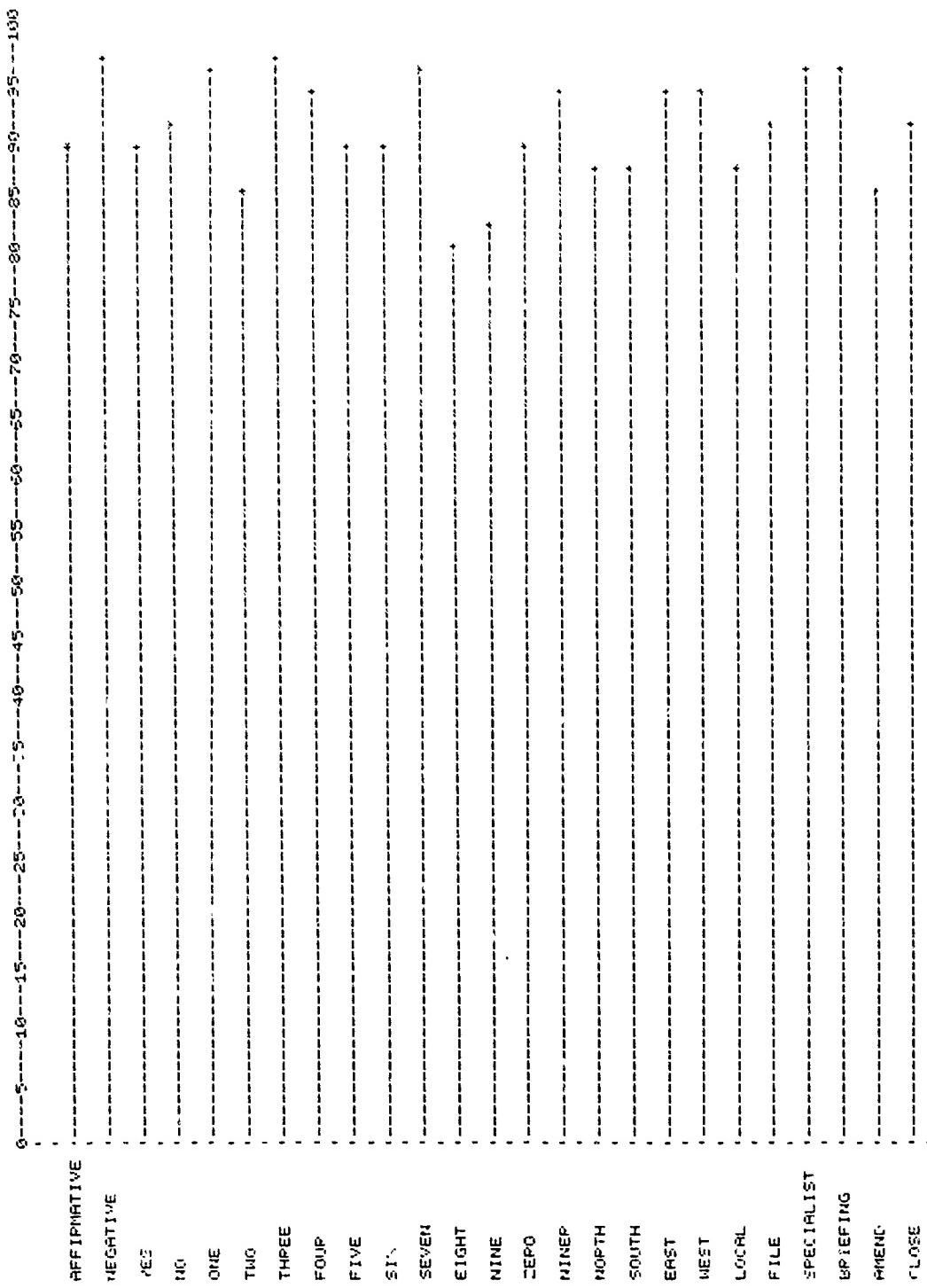
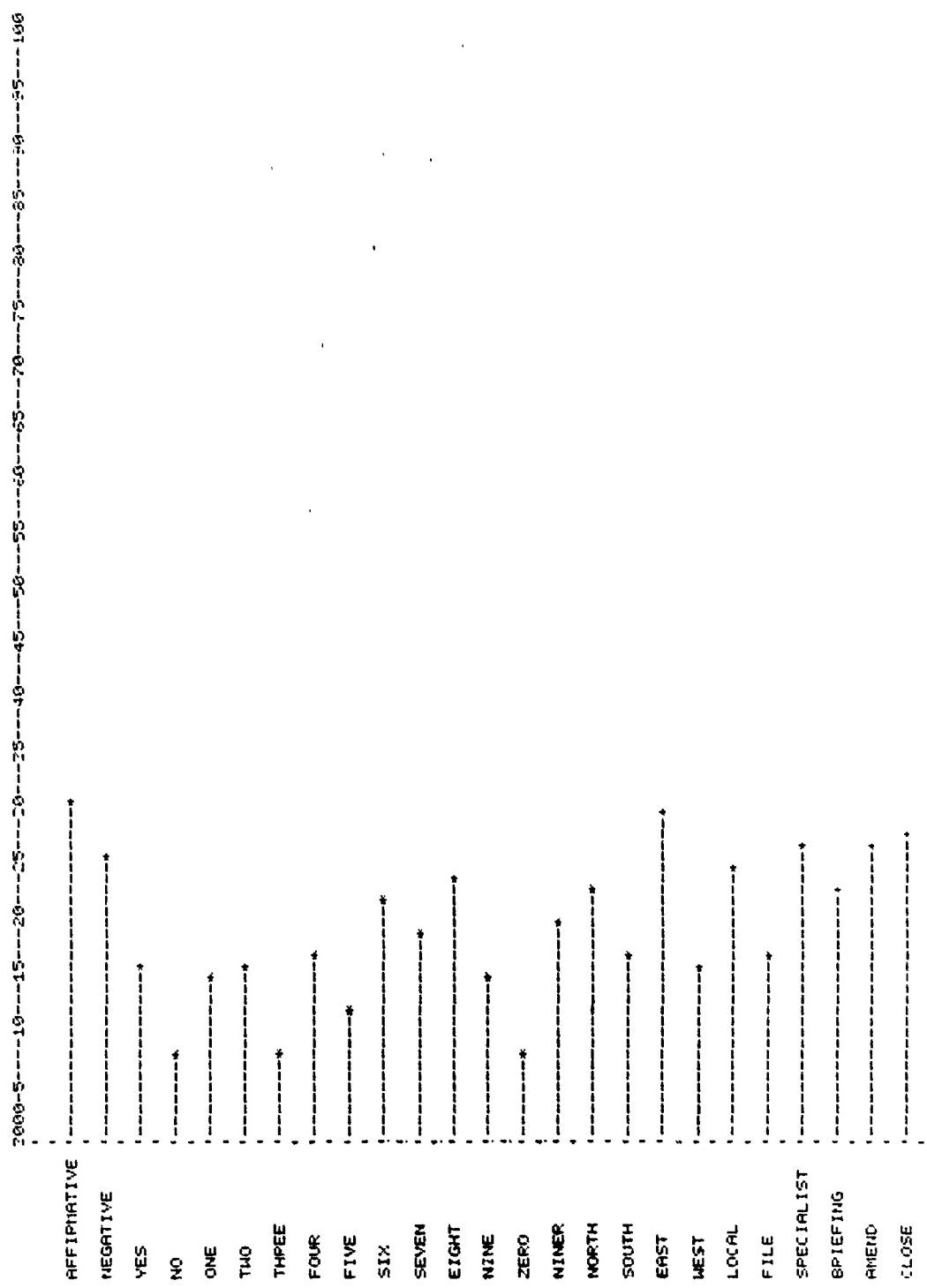


FIGURE A-3C. PILOTS — SIMULATED SUBGROUP RESTRICTIONS

MODIFIED PERCENT CORRECT
80-59-A-3C



1/10 MEAN SCORE OVER 20-35
80-59-A-3D

FIGURE A-3D. PILOTS — COMPARISON OF RELATIVE QUALITY SCORES

FIRST CHOICE DISTRIBUTION

	AFFIRMATIVE	NEGATIVE	YES	NO
AFFIRMATIVE	48	51	52	51
NEGATIVE	6	6	6	6
YES	6	6	6	6
NO	6	6	6	6
ONE	1	1	1	1
TWO	0	0	0	0
THREE	0	0	0	0
FOUR	0	0	0	0
FIVE	0	0	0	0
SIX	0	0	0	0
SEVEN	0	0	0	0
EIGHT	0	0	0	0
NINE	0	0	0	0
ZERO	0	0	0	0
NINEF	0	0	0	0
NCPTH	0	0	0	0
SOUTH	0	0	0	0
EAST	0	0	0	0
WEST	0	0	0	0
LOCAL	0	0	0	0
FILE	0	0	0	0
SPECIALIST	0	0	0	0
BRIEFING	0	0	0	0
AMEND	0	0	0	0
CLOSE	0	0	0	0

	ONE	NO	YES	NO
AFFIRMATIVE	51	52	51	51
NEGATIVE	6	6	6	6
YES	6	6	6	6
NO	6	6	6	6
ONE	1	1	1	1
TWO	0	0	0	0
THREE	0	0	0	0
FOUR	0	0	0	0
FIVE	0	0	0	0
SIX	0	0	0	0
SEVEN	0	0	0	0
EIGHT	0	0	0	0
NINE	0	0	0	0
ZERO	0	0	0	0
NINEF	0	0	0	0
NCPTH	0	0	0	0
SOUTH	0	0	0	0
EAST	0	0	0	0
WEST	0	0	0	0
LOCAL	0	0	0	0
FILE	0	0	0	0
SPECIALIST	0	0	0	0
BRIEFING	0	0	0	0
AMEND	0	0	0	0
CLOSE	0	0	0	0

80-59-A-3E

FIGURE A-3E. PILOTS — FIRST CHOICE WORD DISTRIBUTION (Sheet 1 of 2)

F-1000 FORM 1000		DATA RECORDED AND PROCESSED BY THE FAA TECHNICAL CENTER	
LOCAL	51	TOTALS	51
WEST	0	CLOSE	2
EAST	1	AMEND	0
SOUTH	0	BRIEFING	0
NORTH	6	SPECIALIST	0
NINEP	0	AFFIRMATIVE	0
AFFIRMATIVE	0	NEGATIVE	0
NEGATIVE	0	YES	0
YES	0	NO	0
ONE	0	ONE	0
TWO	0	TWO	0
THREE	0	THREE	0
FOUR	0	FOUR	0
FIVE	0	FIVE	0
SIX	0	SIX	0
SEVEN	0	SEVEN	0
EIGHT	0	EIGHT	0
NINE	0	NINE	0
ZERO	0	ZERO	0
NINER	0	NINER	0
NORTH	0	NORTH	0
SOUTH	0	SOUTH	0
EAST	0	EAST	0
WEST	0	WEST	0
LOCAL	0	LOCAL	0
FILE	0	FILE	0
SPECIALIST	0	SPECIALIST	0
BRIEFING	0	BRIEFING	0
AMEND	0	AMEND	0
CLOSE	0	CLOSE	0

FIGURE A-3E. PILOTS — FIRST CHOICE WORD DISTRIBUTION (Sheet 2 of 2)

**DATA RECORDED AND PROCESSED
BY THE FAA TECHNICAL CENTER**

FEMALES - UPD		VERIFICATION		CONFUSION			
WORD	GARBLE	206	56	56	56	56	56
AFFIRMATIVE	2560	1200	56	56	56	56	56
NEGATIVE	2560	1200	56	56	56	56	56
YES	2560	1200	56	56	56	56	56
NO	2560	1200	56	56	56	56	56
ONE	2560	1200	56	56	56	56	56
TWO	2560	1200	56	56	56	56	56
THREE	2560	1200	56	56	56	56	56
FOUR	2560	1200	56	56	56	56	56
FIVE	2560	1200	56	56	56	56	56
SIX	2560	1200	56	56	56	56	56
SEVEN	2560	1200	56	56	56	56	56
EIGHT	2560	1200	56	56	56	56	56
NINE	2560	1200	56	56	56	56	56
ZERO	2560	1200	56	56	56	56	56
NINEP	2560	1200	56	56	56	56	56
NORTH	2560	1200	56	56	56	56	56
SOUTH	2560	1200	56	56	56	56	56
EAST	2560	1200	56	56	56	56	56
WEST	2560	1200	56	56	56	56	56
LOCAL	2560	1200	56	56	56	56	56
FILE	2560	1200	56	56	56	56	56
SPECIALIST	2560	1200	56	56	56	56	56
BRIEFING	2560	1200	56	56	56	56	56
ARMEND	2560	1200	56	56	56	56	56
CLOSE	2560	1200	56	56	56	56	56
WORD	RIGHT	1100	56	56	56	56	56
AFFIRMATIVE	15	15	56	56	56	56	56
NEGATIVE	15	15	56	56	56	56	56
YES	15	15	56	56	56	56	56
NO	15	15	56	56	56	56	56
ONE	15	15	56	56	56	56	56
TWO	15	15	56	56	56	56	56
THREE	15	15	56	56	56	56	56
FOUR	15	15	56	56	56	56	56
FIVE	15	15	56	56	56	56	56
SIX	15	15	56	56	56	56	56
SEVEN	15	15	56	56	56	56	56
EIGHT	15	15	56	56	56	56	56
NINE	15	15	56	56	56	56	56
ZERO	15	15	56	56	56	56	56
NINEP	15	15	56	56	56	56	56
NORTH	15	15	56	56	56	56	56
SOUTH	15	15	56	56	56	56	56
EAST	15	15	56	56	56	56	56
WEST	15	15	56	56	56	56	56
LOCAL	15	15	56	56	56	56	56
FILE	15	15	56	56	56	56	56
SPECIALIST	15	15	56	56	56	56	56
BRIEFING	15	15	56	56	56	56	56
ARMEND	15	15	56	56	56	56	56
CLOSE	15	15	56	56	56	56	56
TOTALS	423	93	192	1	38	74	64
TOTAL INTERPRETS	> 526						
PERCENT CORRECT	> 82						
PERCENT THAT	> 37						

FIGURE A-4A. FEMALES — COMPUTER ANALYSIS

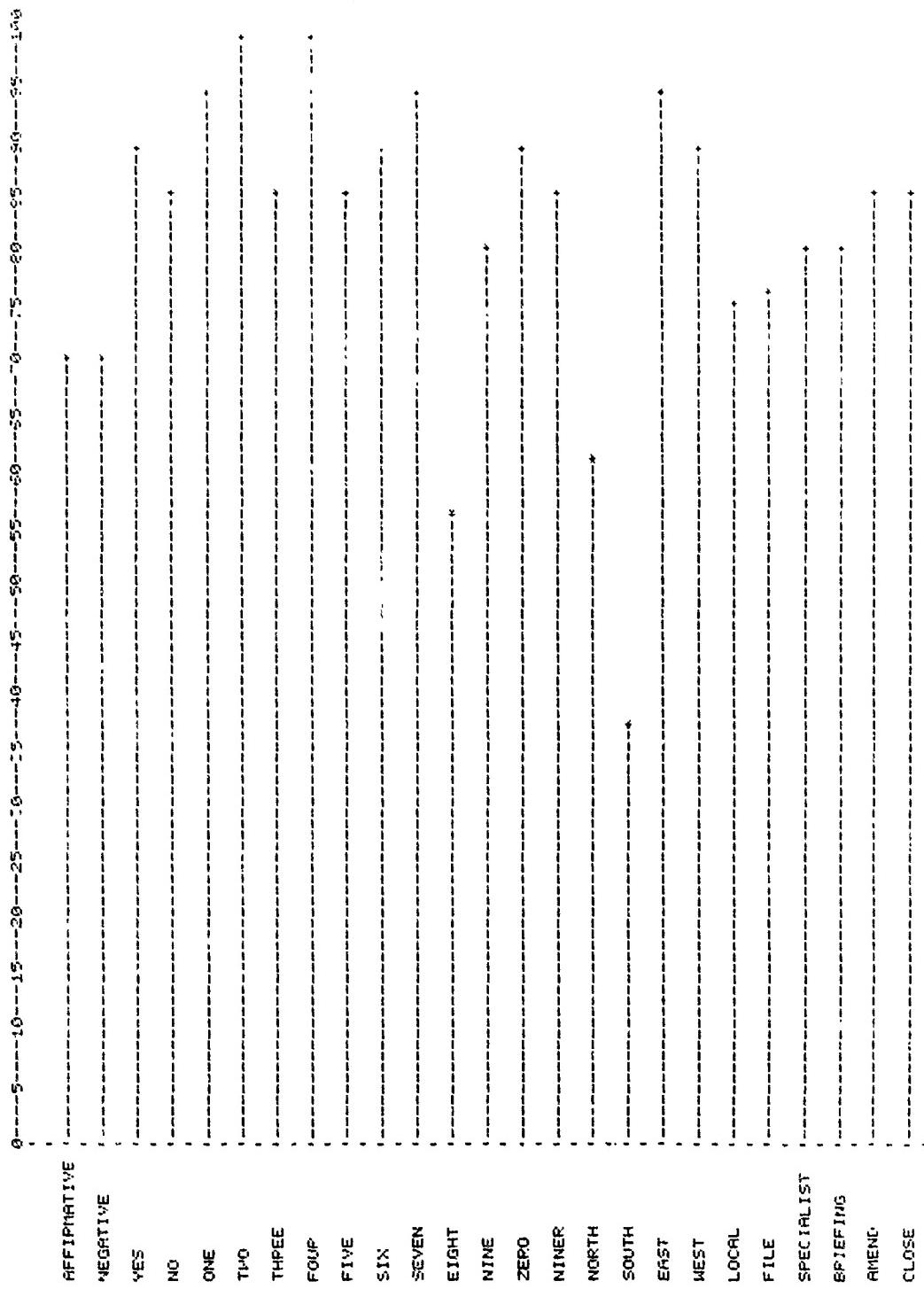
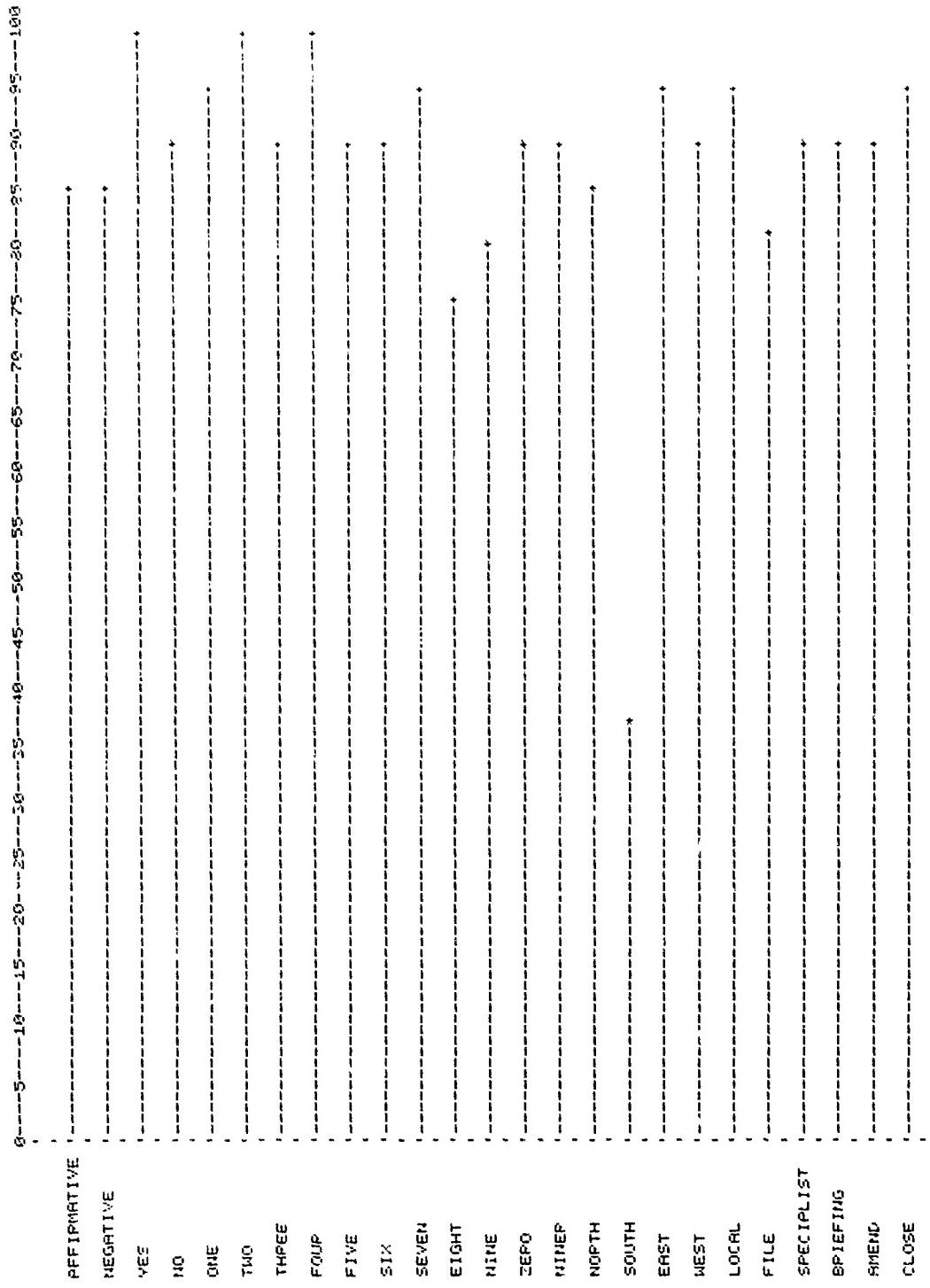


FIGURE A-4B. FEMALES — RELATIVE RECOGNITION RATE

PERCENTAGE OF CORRECT FIRST CHOICE:

30 59-A-48



MODIFIED PERCENT CORRECT

30-59-A-4C

FIGURE A-4C. FEMALES — SIMULATED SUBGROUP RESTRICTIONS

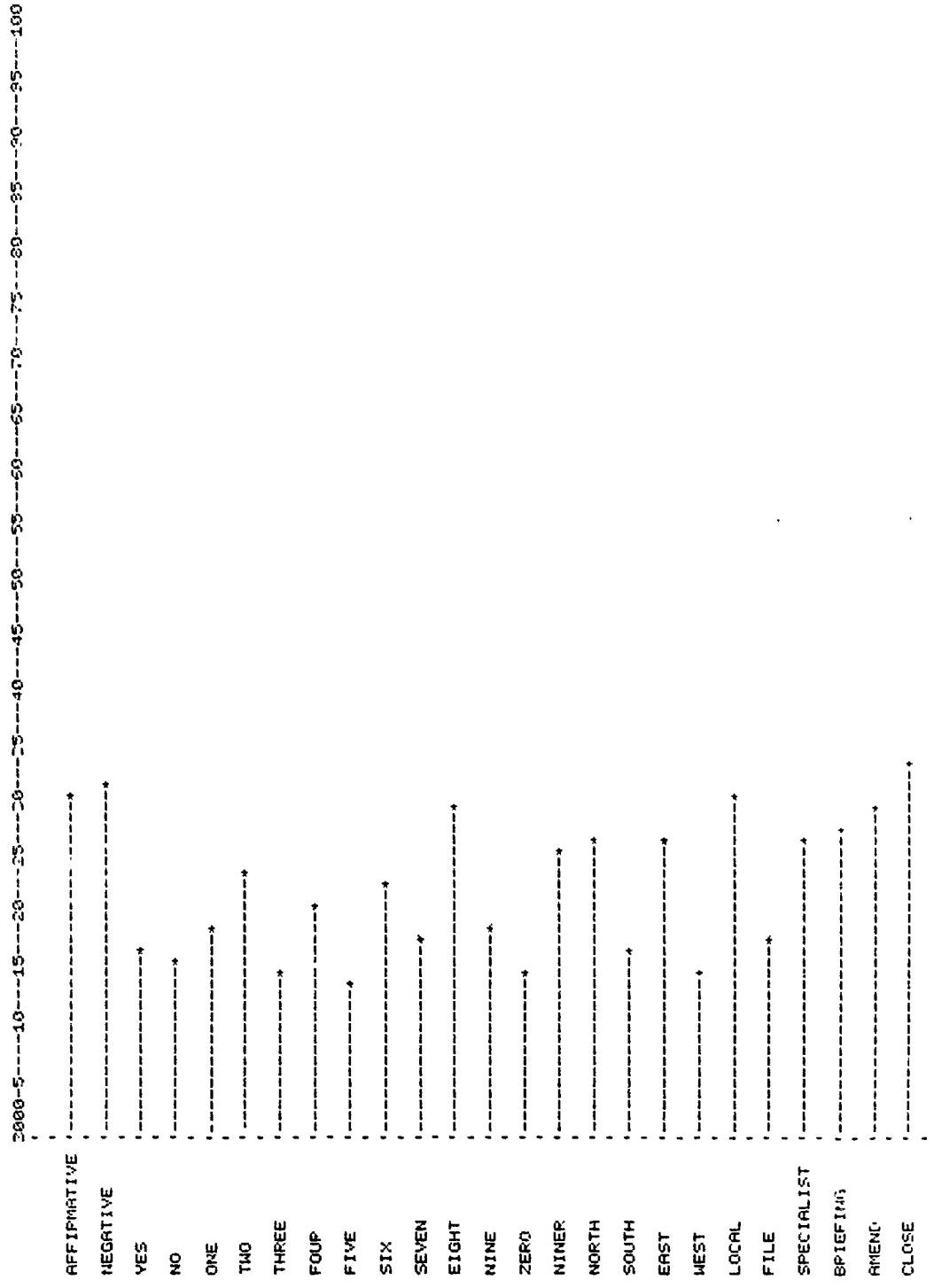


FIGURE A-4D. FEMALES --- CORRELATION OF RELATIVE QUALITY SCORES

FIRST CHOICE DISTRIBUTION

	AFFIRMATIVE	NEGATIVE	YES	NO	TWO	THREE
AFFIRMATIVE	15	0	6	6	0	0
NEGATIVE	0	15	6	6	0	0
YES	0	0	19	1	0	0
NO	0	0	0	18	0	0
ONE	0	0	0	0	0	0
TWO	0	0	0	0	0	0
THREE	0	0	0	0	0	0
FOUR	0	0	0	0	0	0
FIVE	0	0	0	0	0	0
SIX	0	0	0	0	0	0
SEVEN	0	0	0	0	0	0
EIGHT	0	0	0	0	0	0
NINE	0	0	0	0	0	0
ZERO	0	0	0	0	0	0
NUMBER	0	0	0	0	0	0
NORTH	0	0	0	0	0	0
SOUTH	0	0	0	0	0	0
EAST	0	0	0	0	0	0
WEST	0	0	0	0	0	0
LOCAL	0	0	0	0	0	0
FILE	0	0	0	0	0	0
SPECIALIST	0	0	0	0	0	0
BRIEFING	0	0	0	0	0	0
AMEND	0	0	0	0	0	0
CLOSE	0	0	0	0	0	0

FIGURE A-4E. FEMALES — FIRST CHOICE WORD DISTRIBUTION (Sheet 1 of 2)

80-59-A-1E

DATA RECORDED AND PROCESSED
BY THE FAA TECHNICAL CENTER

80-59-4-68

FIGURE A-4E. FEMALES — FIRST CHOICE WORD DISTRIBUTION (Sheet 2 of 2)

APPENDIX B
WORD SCORE DISTRIBUTION

Figure B-1 contains idealized first and second choice word score distribution plots for each vocabulary element. The study of these idealized graphs is useful for developing a subjective concept of the quality of the recognition of each word.

These plots are generated by taking the average score of correct first choice utterances (MEAN) and the average separation from the second choice word (DELTA) and using these values as the maximum points of the first and second choice idealized bell curves. The

actual shape of the curves is derived by plotting the standard deviation values from MEAN (STDM) and DELTA (STDD). Since the plot is an idealization no units are assigned to the Y-axis which is representative of the number of utterances having that quality score.

The reader's attention is directed toward two specific features of each plot. First, the area which lies under both of the word curves and secondly, the proximity of the first choice curve to the default value of VERIFICATION may be considered as being closely related to the number of verification sequences encountered. A detailed explanation of a set of curves may be found in figure 9 of the text.

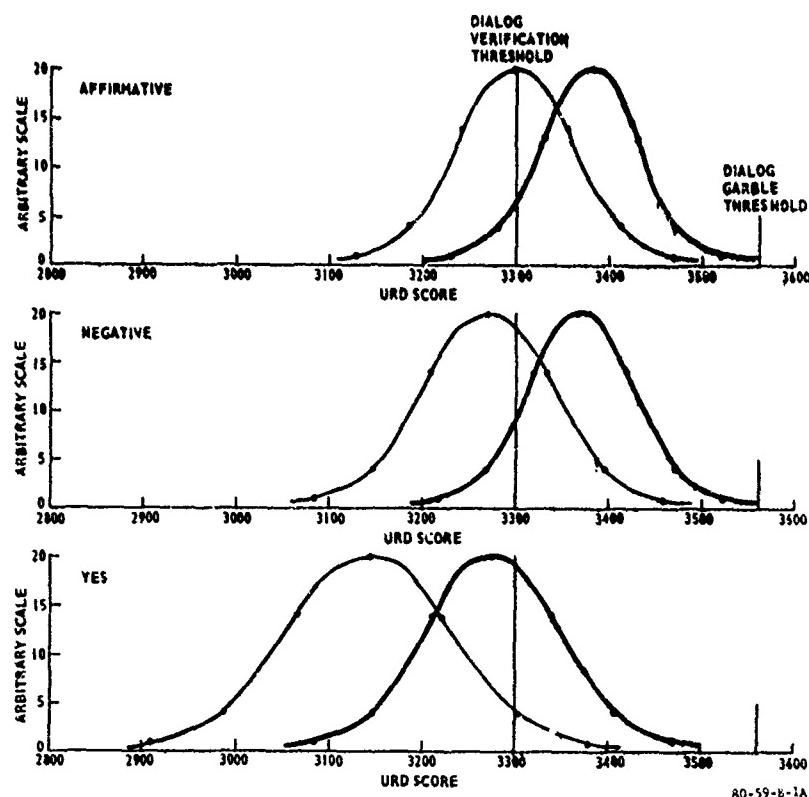


FIGURE B-1. FIRST AND SECOND CHOICE WORD SCORE DISTRIBUTION (Sheet 1 of 5)

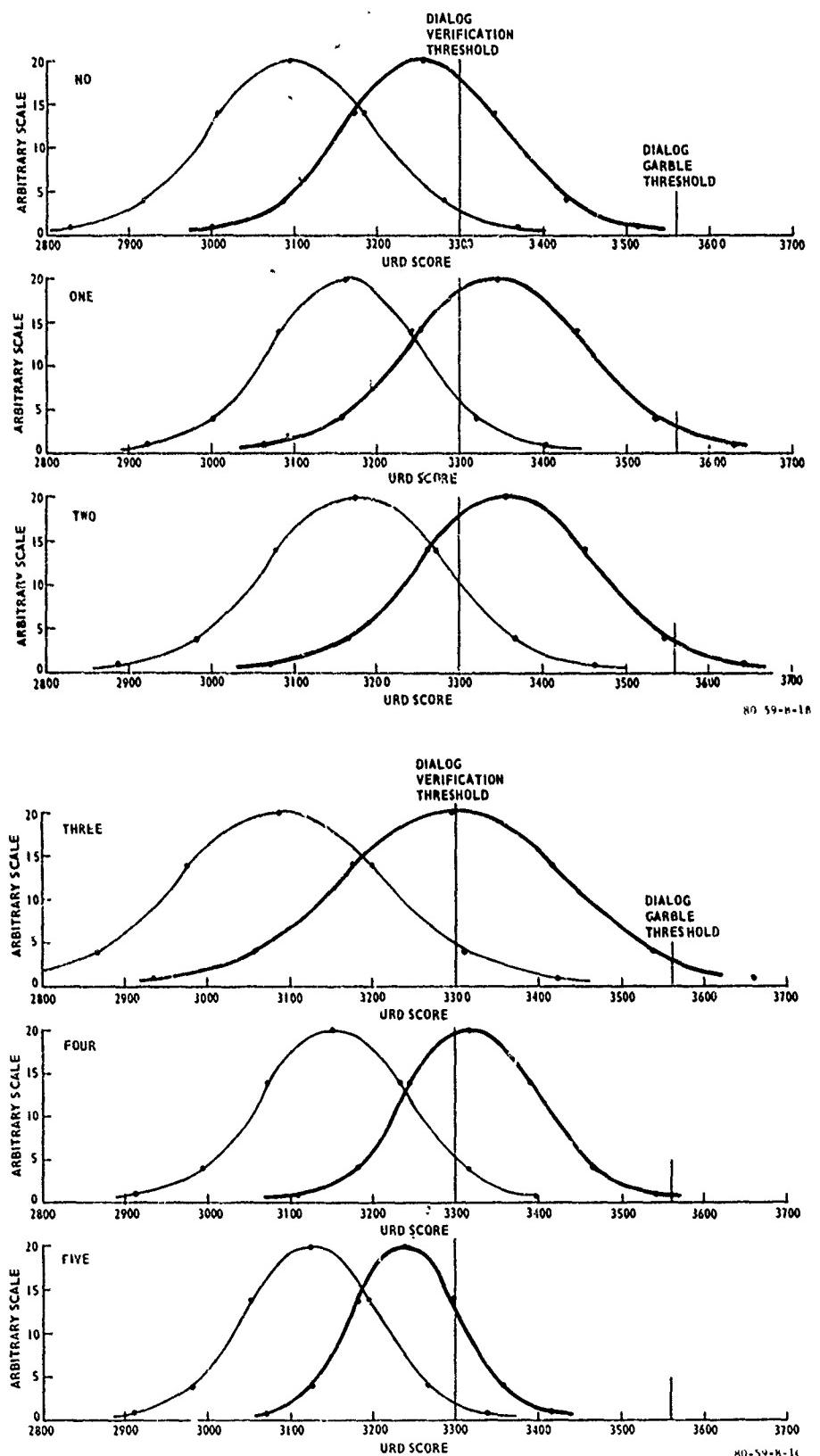


FIGURE B-1. FIRST AND SECOND CHOICE WORD SCORE DISTRIBUTION (Sheet 2 of 5)

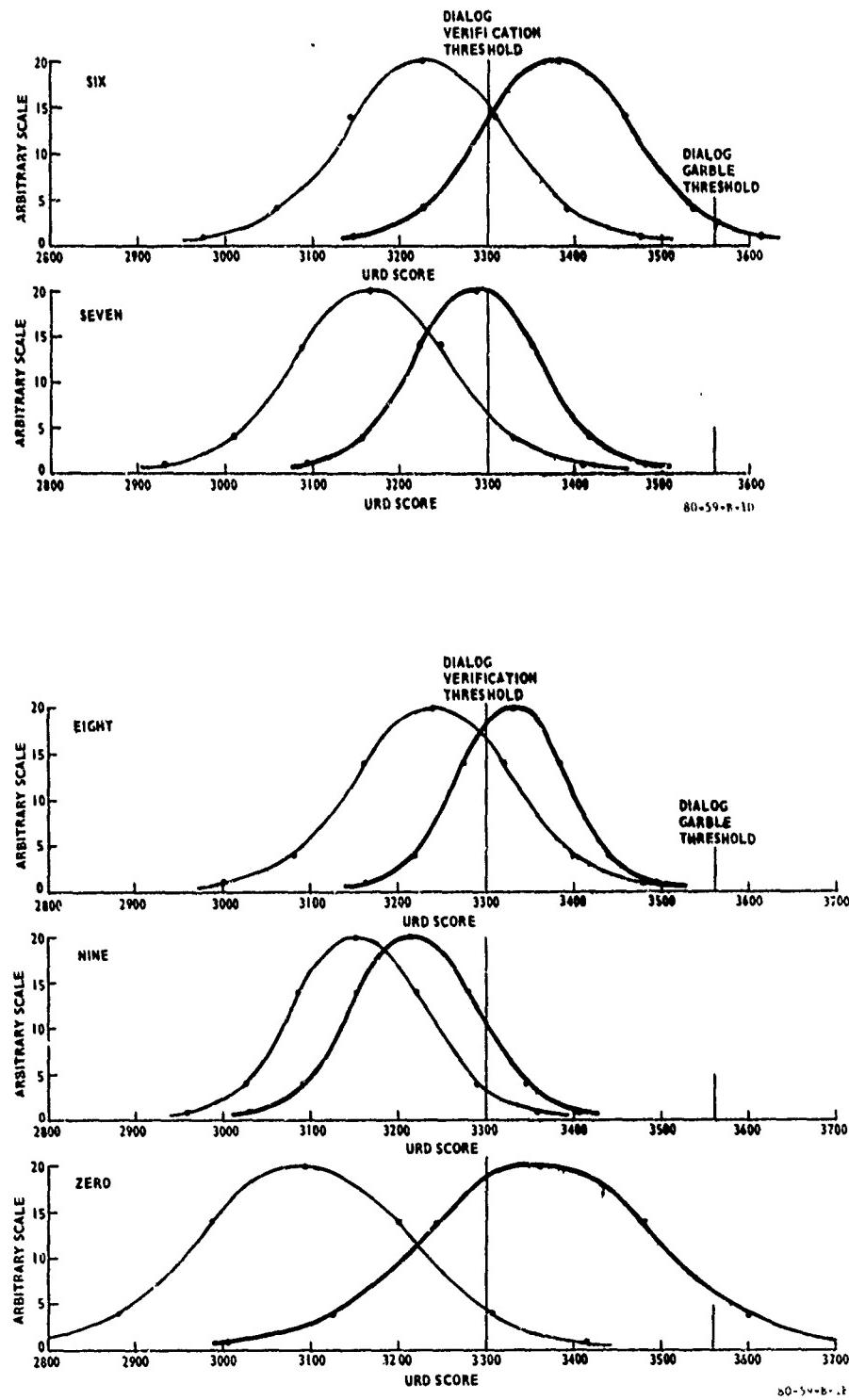


FIGURE B-1. FIRST AND SECOND CHOICE WORD SCORE DISTRIBUTION (Sheet 3 of 5)

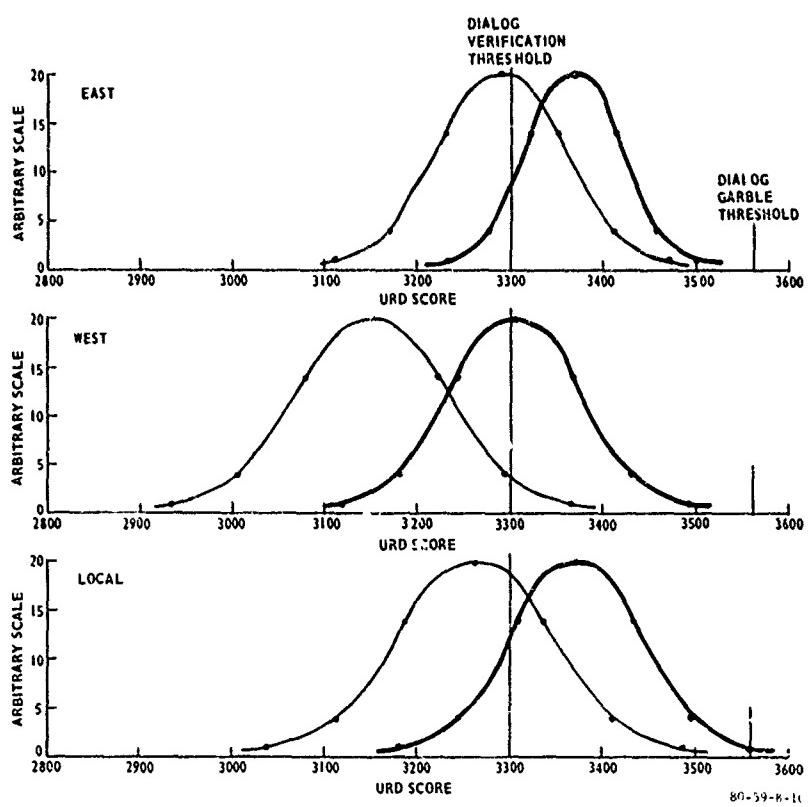
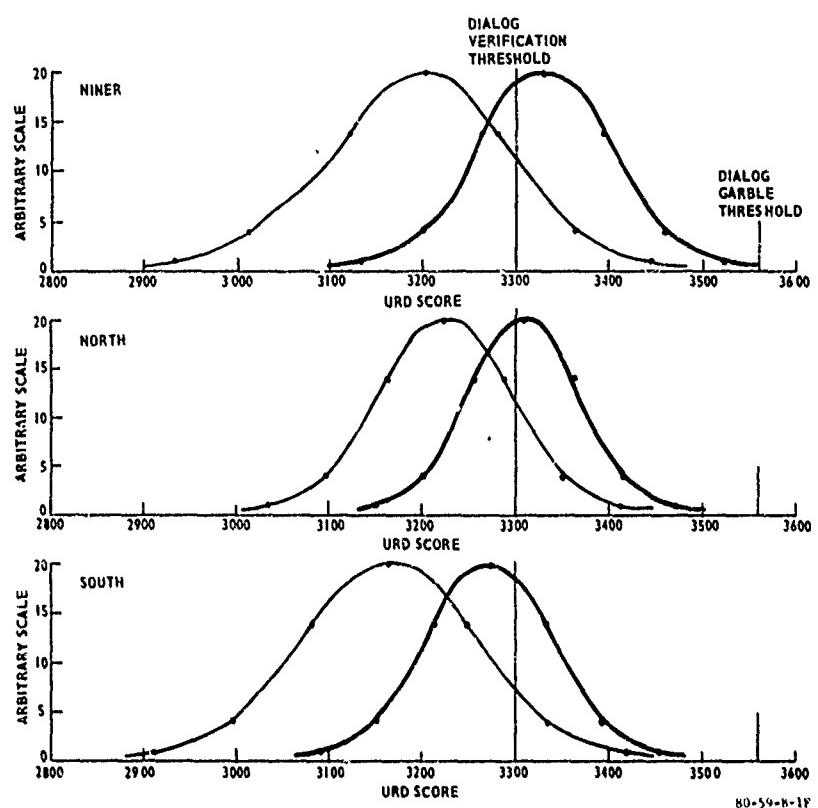


FIGURE B-1. FIRST AND SECOND CHOICE WORD SCORE DISTRIBUTION (Sheet 4 of 5)

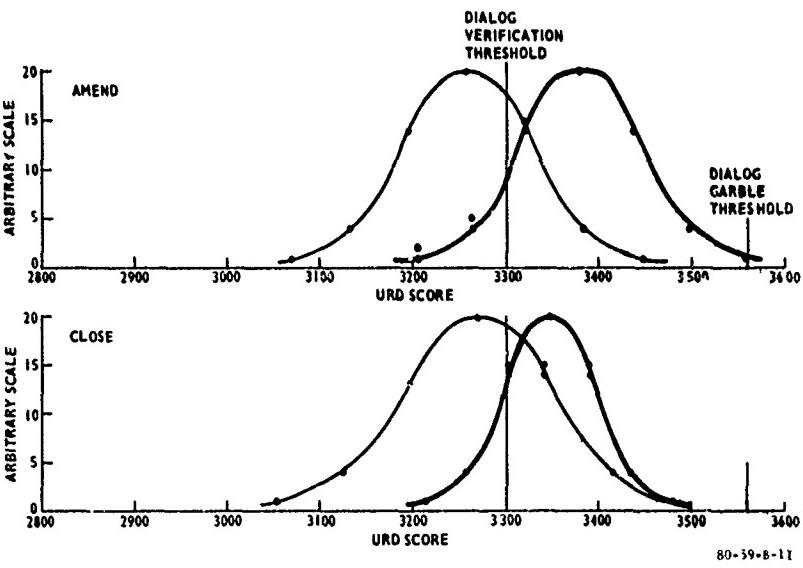
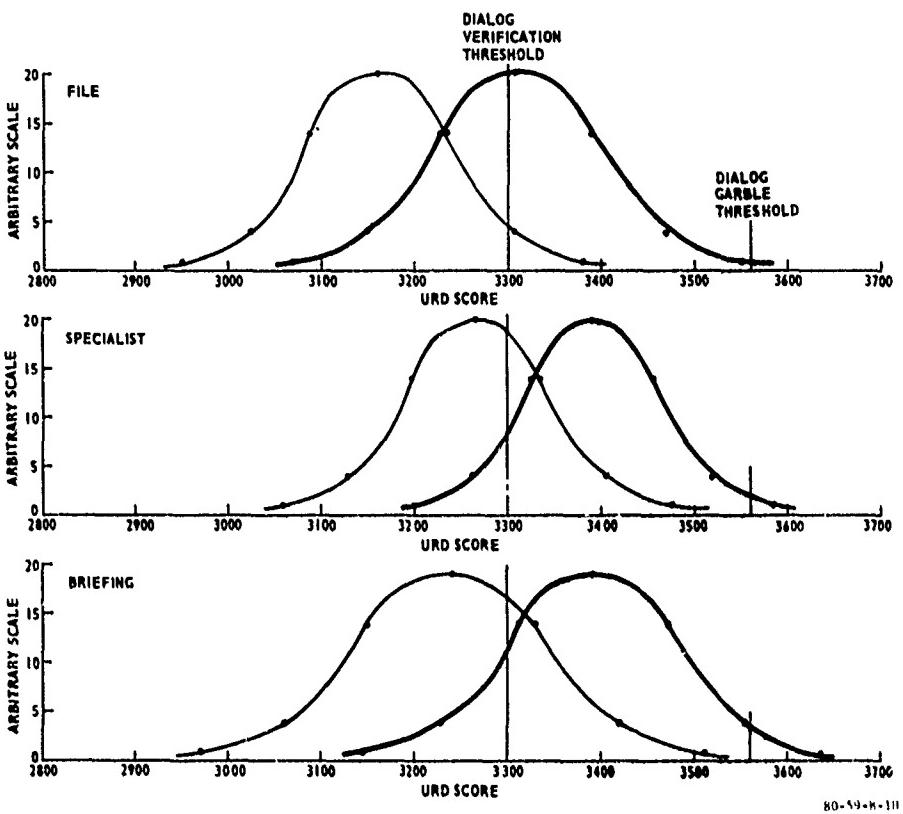


FIGURE B-1. FIRST AND SECOND CHOICE WORD SCORE DISTRIBUTION (Sheet 5 of 5)

APPENDIX C

ALGORITHM EXPLANATION

Appendix C serves to explain and illustrate the algorithm employed to generate the information contained in figure 6 and appendix A. This algorithm was implemented in assembly language on an Interdata 7/32 mini-computer. Numeric quantities were calculated using 32-bit fixed point arithmetic and are rounded to the nearest whole number.

A major assumption was made in the development of the algorithm: all vocabulary elements required for the test would be spoken by the subject and in a given order to enable the program to determine if the first or second choice word was correct. This assumption was forced to be true by the two team members who conducted the test.

Each vocabulary element is assigned an 80-byte-long parameter block containing pertinent information regarding each word as well as reserved storage locations for computed information. The parameter block structure is given in block diagram form in figure C-1.

The first full-word (32 bits) of the parameter block contains a pointer to an area of memory which contains the American Standards Code for Information Interchange (ASCII) equivalent of the word. This segment of the parameter block also serves to correlate a parameter block with its corresponding vocabulary element. For example, the parameter block for the word AFFIRMATIVE begins with a pointer to a memory location labeled AFFIRM.

The third half-word (16 bits) of the parameter block contain the associated words numeric identification code. As an example, the parameter block for the word AFFIRMATIVE would contain 000D, in

bits 32 to 47 (which is the hexadecimal equivalent of 14) the code for AFFIRMATIVE. Table 2 of the report contains a list of word codes.

The next three half-words of the parameter blocks contain the default values of the quality parameters, GARBLE, VERIFICATION, and CONFUSION, in hexadecimal form. These locations may be modified to contain any desired value(s) prior to the beginning of the analysis segment of the program.

The remainder of the parameter block contains zeros at startup time and is used to store computed data. A block diagram of a typical word parameter block follows this description.

The advantage of using a parameter block data structure for each vocabulary element lies in the fact that indexing to appropriate storage locations is greatly simplified. This structure also facilitates the modification of the program to allow for the expansion of the vocabulary.

The URD operating data is read from a disc file in a single pass. Count values are incremented after each reading, if appropriate. Averages and standard deviations are computed after all data records have been read. The standard deviations are computed using the following equation:

$$\text{std.} = \sqrt{\frac{\sum(X^2) - (\sum X)^2}{N - 1}}$$

X = The datum being operated upon.

N = The number of elements under consideration.

Square roots, required for standard deviations, are computed using the Newton-Raphson technique of successive approximations. Sixteen passes are executed to derive each root.

The graphs and distribution analysis are computed after the analysis segment of the routine. Graphs are generated using the appropriate information contained in the word parameter blocks. The data required to generate a first choice word distribution analysis is contained in a separate buffer which

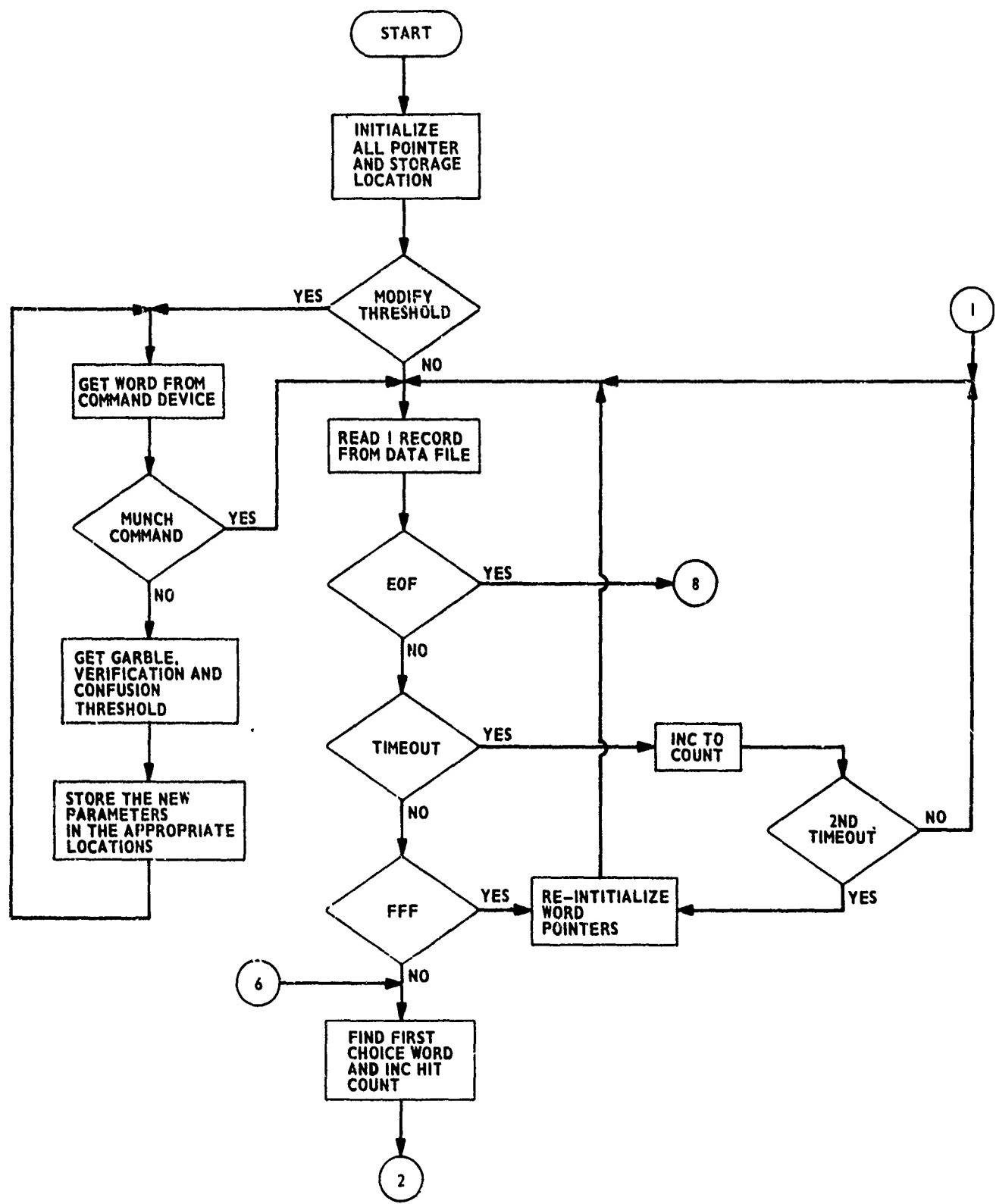
was filled during the analysis phase. Graph and distribution printouts may be inhibited via an operator command if not desired.

Figure C-2 is a macro flow chart of the analysis segment of the algorithm.

POINTER TO ASCII CODE	0	2	
DIALOG WORD CODE	4	6	GARBLE THRESHOLD
VERIFICATION THRESHOLD	8	10	CONFUSION THRESHOLD
DIALOG SUBGROUP	12	14	SUBGROUP WRONG COUNT
RIGHT COUNT	16	18	WRONG COUNT
WAS THAT COUNT	20	22	
TIMEOUT COUNT	24	26	SECOND CHOICE CORRECT COUNT
SUM OF SCORE SEPARATION	28	30	
MOST SIGNIFICANT HALF OF SQUARED SCORE SUMMATION	32	34	
SUM OF THE SQUARE OF THE DELTA VALUES	36	38	
SUM OF CORRECT AMPLITUDES	40	42	
MOST SIGNIFICANT HALF OF THE SQUARED RIGHT AMPLITUDE SUM	44	46	
SUM OF WRONG AMPLITUDES	48	50	
MOST SIGNIFICANT HALF OF THE SQUARED RIGHT AMPLITUDE SUM	52	54	NOTE: MEMORY LOCATION NUMBERS ARE IN BYTES. ALL WORD PARAMETER BLOCKS ARE 80 BYTES LONG
LEAST SIGNIFICANT HALF OF THE SQUARED RIGHT AMPLITUDE SUM	56	58	
SUM OF WRONG AMPLITUDES	60	62	
MOST SIGNIFICANT HALF OF THE SQUARED WRONG AMPLITUDE SUM	64	66	
LEAST SIGNIFICANT HALF OF THE SQUARED WRONG AMPLITUDE SUM	68	70	
STORAGE FOR PERCENT CORRECT	72	74	STORAGE FOR MODIFIED PERCENT CORRECT
NUMBER OF 'WAS THATS' ON WRONG 1ST CHOICES	76	78	NUMBER OF WAS THATS DUE TO INSUFFICIENT DELTA
POINTER TO NEXT WORD ASCII CODE	0	2	

80-59-C-1

FIGURE C-1. WORD PARAMETER BLOCKS



80-59-C-2A

FIGURE C-2. ANALYSIS ALGORITHM FLOWCHART (Sheet 1 of 6)

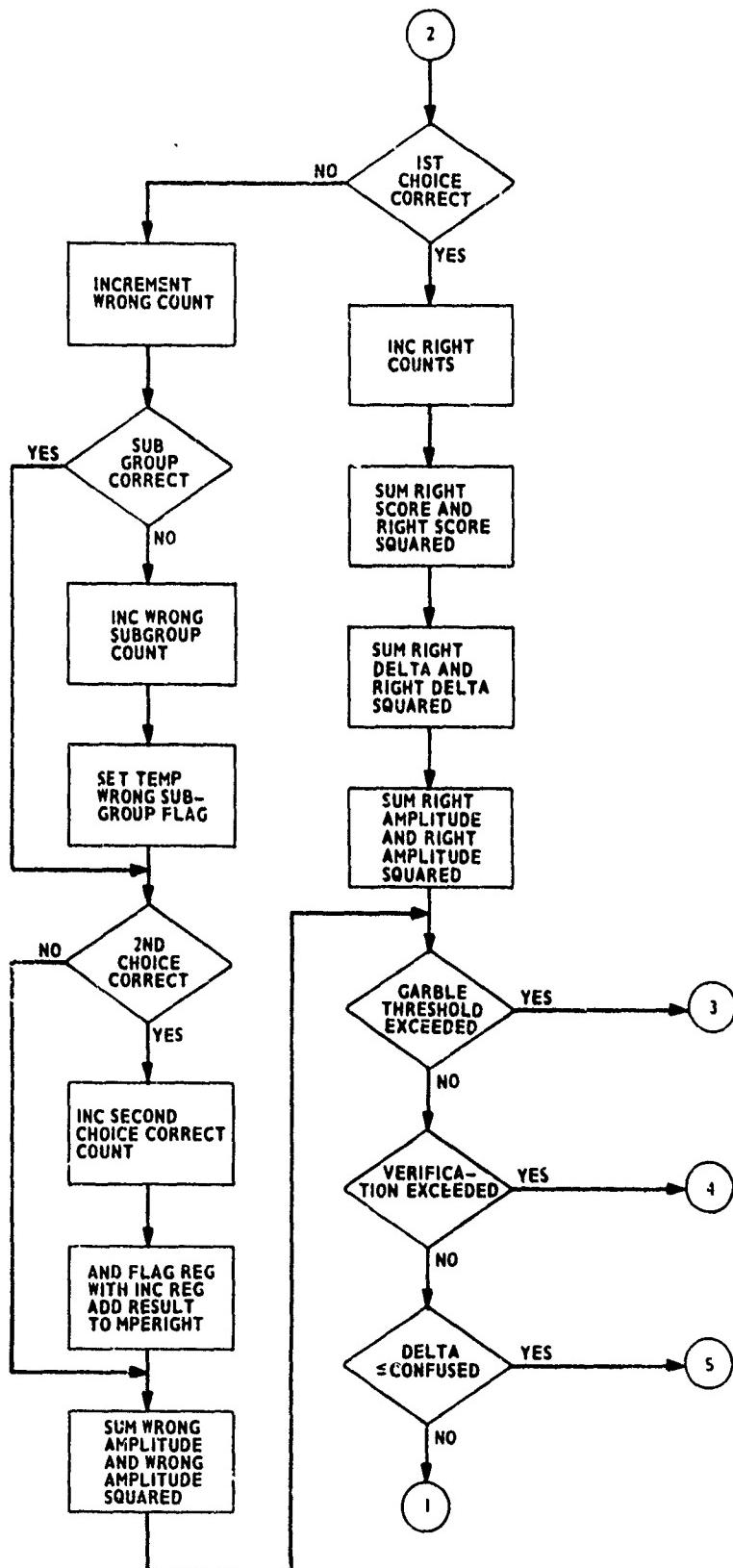


FIGURE C-2. ANALYSIS ALGORITHM FLOWCHART (Sheet 2 of 6)

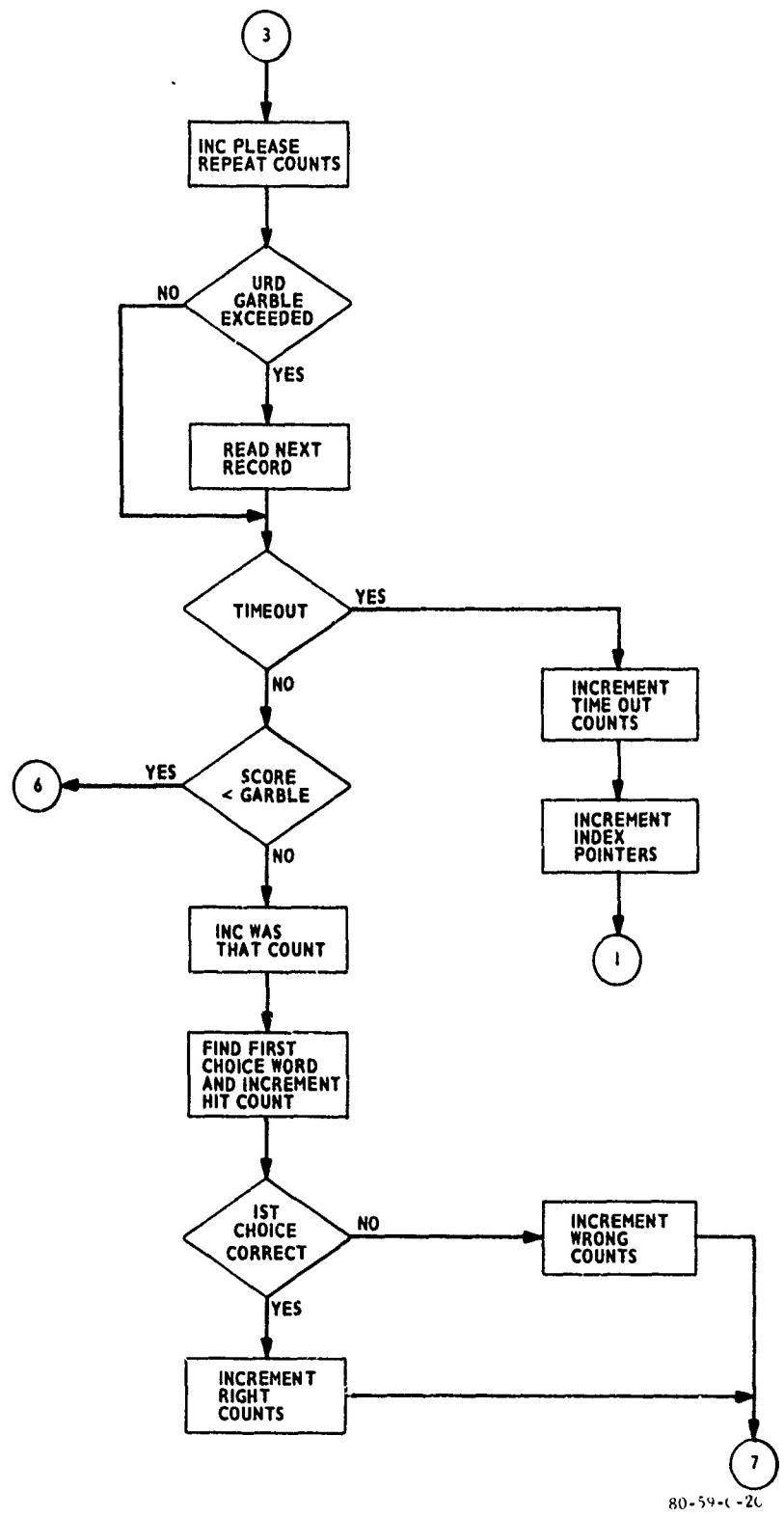
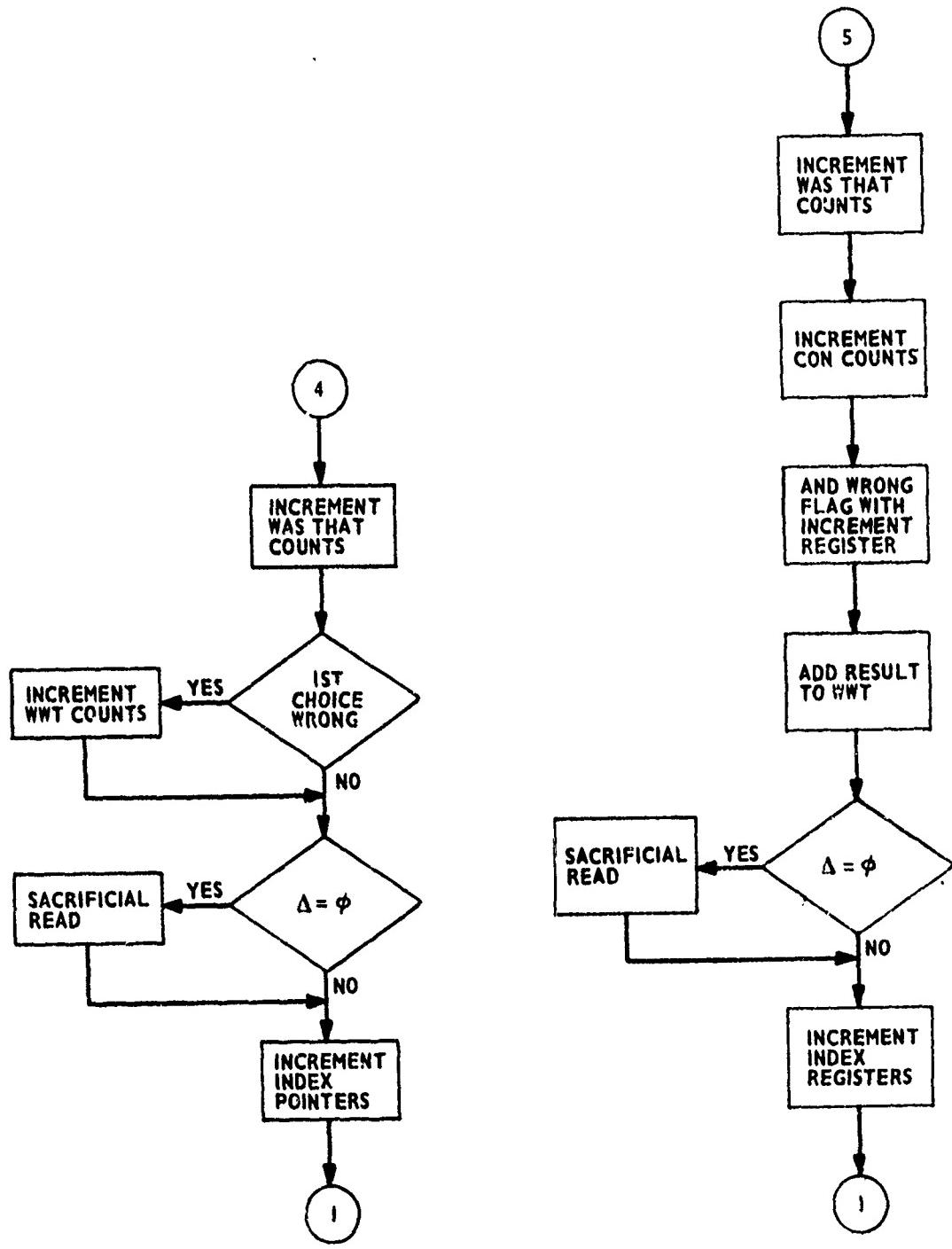


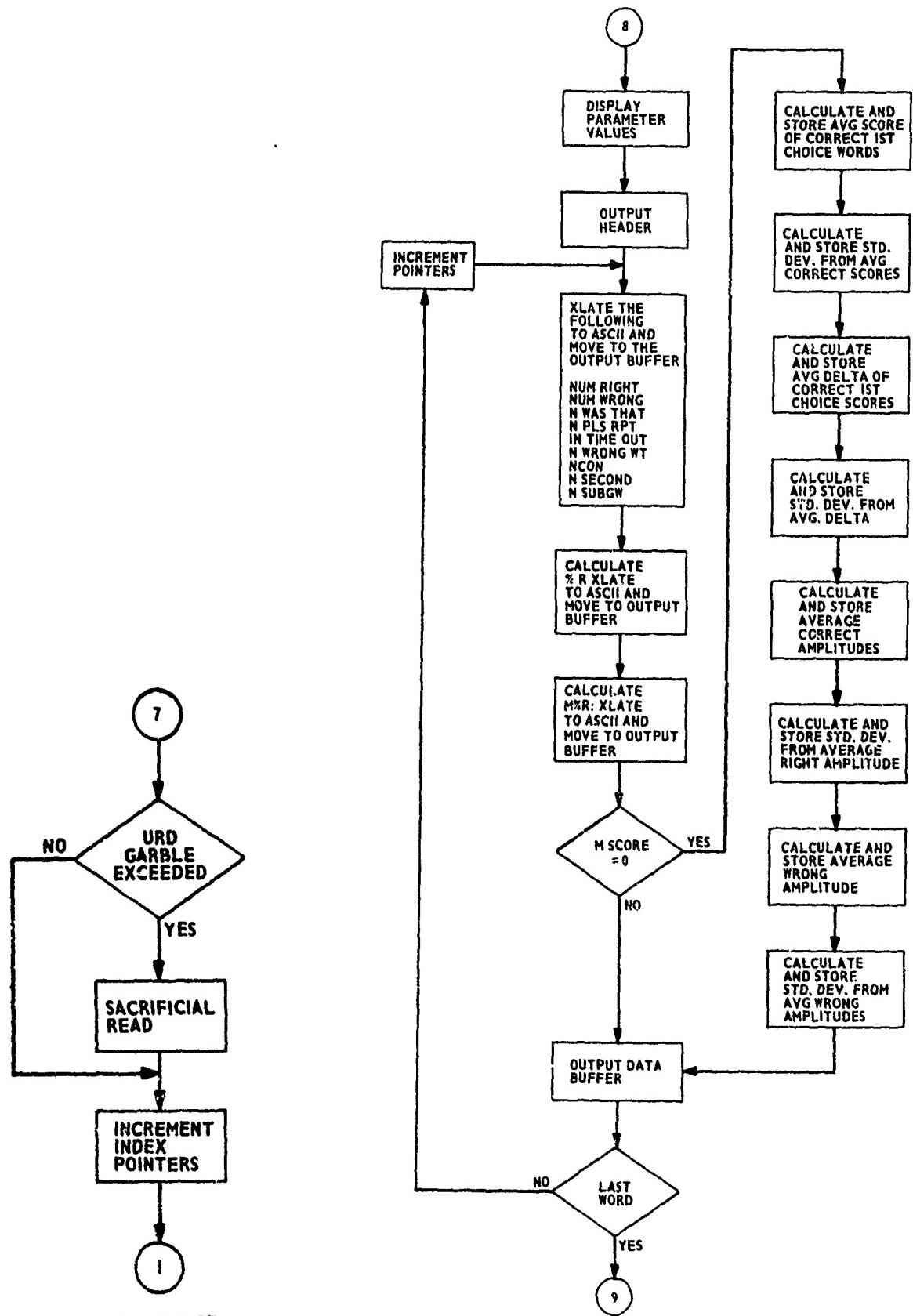
FIGURE C-2. ANALYSIS ALGORITHM FLOWCHART (Sheet 3 of 6)



80-59-C-2D

80-59-C-2E

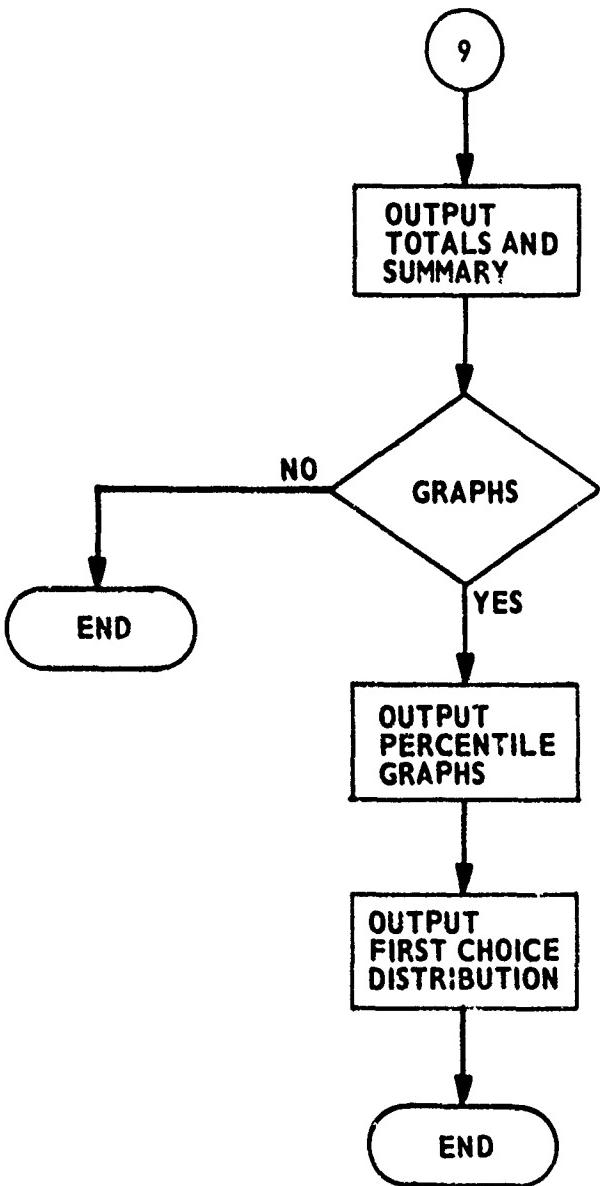
FIGURE C-2. ANALYSIS ALGORITHM FLOWCHART (Sheet 4 of 6)



80-59-C-2F

80-59-C-2G

FIGURE C-2. ANALYSIS ALGORITHM FLOWCHART (Sheet 5 of 6)



80-59-C-2H

FIGURE C-2. ANALYSIS ALGORITHM FLOWCHART (Sheet 6 of 6)